B2

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 December 2000 (28.12.2000)

PCT

(10) International Publication Number
WO 00/78964 A1

(51) International Patent Classification[7]: C12N 15/16, 15/62, 5/10, C07K 14/59, 16/26, A61K 38/22, A01K 67/027, G01N 33/50

(21) International Application Number: PCT/US00/16541

(22) International Filing Date: 15 June 2000 (15.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/334,809 17 June 1999 (17.06.1999) US

(71) Applicant: AMGEN INC. [US/US]; One Amgen Center Drive, Thousand Oaks, CA 91320-1799 (US).

(72) Inventors: PASZTY, Christopher, J.; 1027 Westridge Drive, Ventura, CA 93003 (US). CAO, Jin; 5731 Etiwanda Avenue #5, Tarzana, CA 91356 (US). SCULLY, Sheila; 315 Blue Dolphin Drive, Port Hueneme, CA (US). SCHULTZ, Henry, Joseph; 21827 Parvin Drive, Santa Clarita, CA 91350 (US).

(74) Agents: ODRE, Steven, M. et al.; Amgen, Inc., One Amgen Center Drive, Thousand Oaks, CA 91320-1799 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 00/78964 A1

(54) Titl : GLYCOPROTEIN HORMONE FAMILY MEMBER

(57) Abstract: A novel alpha-like member ("alpha 2") of the human glycoprotein hormone family and nucleic acid molecules encoding it hav been discovered. Vectors, host cells, and methods for producing this alpha 2 polypeptide are disclosed. Also described are methods for the use of alpha 2, including methods for the production of antibodies and for the diagnosis and treatment of disorders associated with alpha 2.

GLYCOPROTEIN HORMONE FAMILY MEMBER

Field of the Invention

The present invention relates to a novel alpha-like member (α2) of the glycoprotein hormone family and nucleic acid molecules encoding same. The invention also relates to vectors, host cells, antibodies and methods for producing alpha 2 polypeptides. Also provided for are methods for the use of alpha 2, including methods for the diagnosis and treatment of disorders associated with alpha 2.

Background of the Invention

Heretofore, there have been five known glycoprotein hormone polypeptides produced in humans: α-subunit, TSH-(thyroid stimulating hormone)-β-subunit, FSH-(follicle stimulating hormone)-β-subunit, LH-(luteinizing hormone)-β-subunit, and CG-(chorionic gonadotropin)-β-subunit; Thotakura and Blithe, Glycobiology, Volume 5, pages 3-10 (1995); Wondisford et al. in Volume 1, Endocrinology (edited by L. DeGroot), pages 208-217, W. B. Saunders Company, Philadelphia, PA (1995); Moyle and Campbell, in Volume 1, Endocrinology (edited by L. DeGroot), pages 230-241, W. B. Saunders Company, Philadelphia, PA (1995). These polypeptides are produced by single genes, with the exception of the CG-β-subunit which is encoded by a multigene cluster composed of six homologous sequences linked to the single LH-β-subunit gene on chromosome 19; Bo and Boime, Journal of Biological Chemistry, vol. 267, pp. 3179-3184 (1992).

Monomeric alpha-subunit (FAS, or free alpha-subunit) has hormonal activity and is secreted by the

pituitary gland and the placenta. FAS has been found to play a role in the differentiation of prolactin producing cells in the pituitary and placenta; see Begeot et al., *Science*, vol. 226, pp. 566-568 (1984), Van-Bael and Denef, *Journal of Neuroendocrinology*, vol. 8, pp. 99-102 (1996), and Moy et al., *Endocrinology*, vol. 137, pp. 1332-1339 (1996); and also to stimulate placental prolactin secretion; see Blithe et al., *Endocrinology*, vol. 129, pp. 2257-2259 (1991).

Alpha-subunit also heterodimerizes with each of the four beta-subunits to form four heterodimeric hormones (TSH, FSH, LH and CG). TSH, FSH and LH are produced in the pituitary, stored in secretion granules, and secreted when the appropriate releasing hormone is produced by the hypothalamus. CG is produced in the placenta and appears to be secreted constitutively (it is not stored in secretion granules); see Wondisford et al. in Volume 1, *Endocrinology* (ed. L. DeGroot), pp. 208-217, above, and Hall and Crowley, Jr. in Volume 1, *Endocrinology* (ed. L. DeGroot), pp. 242-258, W. B. Saunders Company, Philadelphia, PA (1995).

TSH influences basal metabolism by regulating the production of thyroid hormones and is used clinically for enhancing the detection and treatment of thyroid carcinoma; see McEvoy, G.(ed.), *AHFS Drug Information*, pp. 2041-2042, American Society of Health-System Pharmacists, Inc., Bethesda, MD (1998). In addition, diagnostic tests for measuring TSH levels in the blood are commonly used for determining the functional status of the thyroid gland when thyroid gland disorder is suspected.

FSH and LH play important roles in the maintenance of reproductive function in males and females (i.e., gonadal maturation and gonadal steroid production). CG is involved in the maintenance of pregnancy by stimulating the corpus luteum to produce steroid hormones during the first trimester. FSH, LH and CG are used clinically to treat infertility and also as reagents in assisted reproduction procedures such as *in vitro* fertilization (IVF); see McEvoy, G.(ed.), *AHFS Drug Information*, pp. 2564-2567, American Society of Health-System Pharmacists, Inc., Bethesda, MD (1998). Diagnostic tests for measuring FSH, LH and CG levels are used for the diagnosis of fertility disorders, as well as to test for pregnancy.

Naturally occurring metabolites of the above mentioned glycoprotein hormone polypeptides have been described, such as the β-core fragment which is derived from the beta subunit of CG, but no function has yet been assigned to these metabolites; Moyle and Campbell in Volume 1 *Endocrinology* (ed. L. DeGroot) pp. 230-241, above.

In 1994, the five glycoprotein hormone polypeptides were placed into the cystine-knot growth factor structural superfamily, based on the crystal structure of human CG; Lapthorn et al., *Nature*, vol. 369, pp. 455-61 (1994). This superfamily includes the TGF-β (transforming growth factor beta), NGF (nerve growth factor) and PDGF (platelet-derived growth factor) gene families. The cystine-knot is formed by three intramolecular disulfide bonds, has a very characteristic structure, and is responsible for the overall three-dimensional structure of all of the members of the superfamily; Isaacs, *Current Opinion in Structural Biology*, vol. 5, pp. 391-395 (1995).

Summary of the Invention

The present invention provides, in part, an isolated secretable polypeptide, from each of two mammalian species (human and mouse), which is a novel alpha-like member of the glycoprotein hormone family and is herein designated as "alpha 2" or "α2". The invention also provides isolated nucleic acid molecules encoding such polypeptides.

An example of the human amino acid sequence, which includes the "signal" peptide responsible for secretion through the host cell membrane, is that of SEQ ID NO: 1. An example of the mouse amino acid sequence (including signal peptide region) is that of SEQ ID NO: 2. Nucleic acid molecules encoding the polypeptides are of SEQ ID NO: 1 (human) and SEQ ID NO: 2 (mouse) are those of SEQ ID NO: 3 and SEQ ID NO: 4, respectively, among others which are described further below.

As used throughout this disclosure, the terms "gene", "DNA", and "nucleic acid molecule" are meant to refer to isolated molecules which are free of total genomic DNA of a particular species, while retaining coding sequences for the described polypeptides.

Full length amino acid sequences of human and mouse polypeptides in accordance with this invention are shown in Figure 1. N-terminal signal peptides predicted for both of these particular proteins are shown underlined. Both the human and mouse polypeptides contain two asparagine (N) residues located within classic NxT glycosylation motifs (where x denotes any amino acid except for proline and T denotes threonine). These asparagines, which are very

likely to be glycosylated, are located at positions 37 and 81 in the human amino acid sequence (SEQ ID NO: 1) and at positions 36 and 80 in the mouse amino acid sequence (SEQ ID NO: 2).

The signal peptide cleavage site in the human amino acid sequence is expected to be within the region of seven amino acids shown boxed in Figure 1. Thus, the possible "mature" forms for the human protein are represented by the amino acid sequences of SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11 and SEQ ID NO: 12. Signal peptide cleavage at the site which is most likely to be the authentic *in vivo* cleavage site is represented by SEQ ID NO: 8. For the mouse polypeptide the signal peptide cleavage site is expected to be within the region of eight amino acids shown boxed in Figure 1. The high degree of homology (84% identity and 85% similarity) between the amino acid sequences encoded by the human and mouse genes, shown in Figure 2, demonstrates that these genes are orthologs of one another. This novel protein, from human and from mouse, is related to the alpha polypeptide of the glycoprotein hormone family.

The full coding regions of the human and mouse polypeptides of this invention were identified using a combination of cDNA clone sequencing, 5' RACE and Northern analysis. The starting point for these experiments was a public database mouse EST (GenBank accession no. AA709641) that had low but potentially significant homology to a glycoprotein hormone alpha polypeptide amino acid profile that had been created in house. Nothing in the GenBank report for AA709641 identified or suggested a full coding region, nor a significant homology to any known gene.

To determine if the human polypeptide of this invention was related to any known polypeptides, the most likely "mature" form (i.e, processed *in situ* to remove the signal peptide, SEQ ID NO: 8) was run against the NonRedundant Protein database using the program BLAST to search for homologies (specifically, commonly occurring, or "conserved", amino acid residues). Ten of the top fifteen hits were to the glycoprotein hormone alpha polypeptide from various species. GAP analysis indicated that the homology to human glycoprotein hormone alpha polypeptide was at the level of 27% identity and 34% similarity (See Figure 3A). A more careful analysis of this homology has revealed that the identical (conserved) amino acids include those critical for the structure and function of the human glycoprotein hormone alpha polypeptide (See Figure 3B). Additionally, the genomic structure of the human gene that encodes the polypeptide of this invention is virtually identical to that of the human glycoprotein hormone alpha polypeptide gene (See Figure 4). These comparisons are described in greater detail in the section below entitled "Detailed Description of the Invention". In light of these important structural and functional identities, the proteins of the present invention are being referred to herein as glycoprotein hormone "alpha 2" or "α2" polypeptide, as they represent a new alpha-like member of the glycoprotein hormone family.

In greater detail, the polypeptides on this invention comprise the amino acid sequence selected from the group consisting of:

a) the amino acid sequence forth in SEQ ID NO: 1, 5, 6, 7, 8, 9, 10, 11 or 12, optionally further comprising an amino-terminal methionine;

b) a fragment of SEQ ID NO: 1, 5, 6, 7, 8, 9, 10, 11 or 12 comprising about fifty or more amino acid residues, wherein the fragment has an activity of the polypeptide of SEQ ID NOS: 1 and 5-12;

c) an ortholog of SEQ ID NO: 1, 5, 6, 7, 8, 9, 10, 11 or 12; and

d) an analog of (a), (b) or (c) which is at least eighty percent, more preferably at least ninety, and most preferably at least ninety five percent identical, at the amino acid level.

In addition to the aforementioned α2 polypeptides, the present invention also encompasses isolated nucleic acid molecules encoding the polypeptides, as mentioned. Such nucleic acid molecules will comprise:

a) the nucleotide sequence as set forth in SEQ ID NOS: 3, 4, 13, 14, 15, 16, 17, 18, 19 and 20;

b) the nucleotide sequence as set forth in SEQ ID NO: 21;

c) a nucleotide sequence encoding the polypeptide as set forth in SEQ ID NOS: 1, 2, and 5-12;

d) a nucleotide sequence encoding a polypeptide that is at least eighty, more preferably at least ninety, and most preferably at least ninety five percent identical to the polypeptide as set forth in SEQ ID NOS: 5-12 wherein the polypeptide has an activity of alpha 2;

e) an allelic variant or splice variant of any of (a), (b), (c) or (d);

f) a nucleotide sequence which hybridizes under stringent conditions to the complement of any of (a) - (e); and

g) a nucleotide sequence complementary to any of (a), (b), (c), (d), (e) or (f).

The invention also provides for an expression vector comprising the nucleic acid molecules as set forth above, host cells comprising the expression vectors of the invention, and a method of production of an alpha 2 polypeptide comprising culturing the host cells and optionally isolating the polypeptide so produced.

A transgenic non-human animal comprising a nucleic acid molecule encoding an alpha 2 polypeptide is also encompassed by the invention. The alpha 2 nucleic acid molecules are introduced into the animal in a manner that allows expression and increased levels of an alpha 2 polypeptide, which may include increased circulating levels. The transgenic non-human animal is preferably a mammal, and more preferably a rodent, such as a rat or a mouse.

Also provided are derivatives of the alpha 2 polypeptides of the invention, fusion polypeptides comprising the alpha 2 polypeptides of the invention, and antibodies specifically binding the alpha 2 polypeptides of the invention.

Compositions comprising the nucleotides or polypeptides of the invention and a carrier, adjuvant, solubilizer, stabilizer or anti-oxidant or other pharmaceutically acceptable agent are also encompassed by the invention. The compositions may include pharmaceutical compositions comprising therapeutically effective amounts of the nucleotides or polypeptides of the invention, and methods of using the polypeptides and nucleic acid molecules.

Based on the logical inclusion of the $\alpha 2$ polypeptide of this invention in the glycoprotein

hormone family, this polypeptide theoretically could be secreted as a monomer (analogous to FAS) and/or as a heterodimer with one or more beta, or "β", subunits (analogous to TSH, FSH, LH, CG). It has been determined that α2 is expressed in multiple organs, including the pituitary gland and the placenta where the known glycoprotein hormone polypeptides are expressed. This raises the possibility that α2 forms heterodimers with one or more of these polypeptides. Thus, possible heterodimers containing the alpha-2 polypeptide of this invention would be α2/TSH-β, α2/FSH-β, α2/LH-β, α2/CG-β and α2/α. The α2 polypeptide could also form heterodimers with polypeptides which are distinct from the known glycoprotein hormone polypeptides. Within the cystine-knot growth factor superfamily there are many examples of hormones/growth factors which exist as homodimers. Thus, it is also conceivable that α2 is secreted as a homodimer (i.e, α2/α2). Moreover, based on the aforementioned homology between the α and α2 glycoprotein hormone polypeptides, the existence of five distinct α polypeptide-based glycoprotein hormones (FAS, TSH, FSH, LH, CG), and the fact that α2 is expressed in multiple tissues, it is possible that α2 polypeptide forms more than one hormone.

This invention also relates to the use of the α2 polypeptide in co-transfection experiments with each of the five known glycoprotein hormone polypeptides (α, TSH-β, FSH-β, LH-β, and CG-β) to assess possible formation of heterodimers, as described in greater detail in the "Detailed Description of the Invention". The formation of heterodimers for any given pair would indicate the potential for such a heterodimer to exist *in vivo* and would represent a possible α2 heterodimeric hormone. Recombinant α2 heterodimeric hormone could be produced

in mammalian cells, injected into a mammal (for example, mice) and assessed for *in vivo* function and utility in the diagnosis and/or treatment of one or more of the diseases or disorders listed below.

Based on the known functions of the glycoprotein hormones FAS, TSH, FSH, LH, and CG, as well as on the functions of the tissues and organs in which α2 gene expression occurs (see "Detailed Description of the Invention"), α2 polypeptide may be useful, in the form of a monomer or homodimer or as part of a heterodimer, for the treatment of one or more of the following diseases or disorders: obesity, wasting syndromes (for example, cancer associated cachexia), gastrointestinal disorders (for example, ulcers), diabetes, hypertension, immune system dysfunction [for example, excessive inflammation, susceptibility to infection (such as AIDS), poor wound healing, asthma, arthritis and allergies], shock, tissue damage/degeneration (such as caused by cancer treatments), cancers, hyperplasias/hypertrophies, infertility, fertility (contraception) and impotence.

Additionally, this invention concerns the use of nucleic acids (for example, cDNAs) which are homologous to the human α2 cDNA sequence (e.g., SEQ ID NO: 21), or a portion thereof, as a diagnostic reagent for the diseases and disorders mentioned above. Examples of such diagnostic approaches include RT-PCR and nucleic acid hybridization.

This invention is also directed to the utilization of α2 polypeptide for generating and/or identifying antibodies or other alpha 2 binding peptides, such as α2 polypeptide binding proteins (for example, in serum) or random peptides isolated from combinatorial peptide libraries (for example, via

phage display), which recognize and/or bind to regions of the amino acid sequence thereof. This includes subfragments of such binding antibodies and binding peptides. The invention is further concerned with the use of such antibodies and binding peptides in the diagnosis and/or treatment of one or more of the diseases and disorders listed above. Such antibodies and binding peptides would also be very useful for purifying and characterizing α2 polypeptide, α2 polypeptide metabolites and/or complexes containing α2 polypeptide from various natural sources (organs, serum, urine). Recombinant versions of any such forms of α2 polypeptide could be produced in mammalian cells, injected into mice or some other mammal, and assessed for *in vivo* function and utility in the diagnosis and/or treatment of one or more of the diseases or disorders listed above.

In general, binding peptides for alpha 2 as mentioned above will be useful for the same diagnostic and therapeutic applications as detailed herein for antibodies.

The invention also provides for a method of identifying a test molecule which binds to an alpha 2 polypeptide, wherein the method comprises contacting an alpha 2 polypeptide with a test molecule and determining the extent of binding of the test molecule to the polypeptide. The method further comprises determining whether such test molecules are agonists or antagonists of an alpha 2 polypeptide.

The invention also provides for a method of testing the impact of molecules on the expression of alpha 2 polypeptide or on the activity of alpha 2 polypeptide.

A method of regulating expression and modulating (i.e., increasing or decreasing) levels of an alpha 2 polypeptide are also encompassed by the invention. One method comprises administering to a mammal a nucleic acid molecule encoding an alpha 2 polypeptide. In another method, a nucleic acid molecule comprising elements that regulate expression of an alpha 2 polypeptide may be administered. Examples of these methods include gene therapy and anti-sense therapy.

Additionally, because the receptor or receptors for α2 are at present unknown, the α2 polypeptide can be used for identifying receptors thereof, which constitutes still another aspect of this invention. Various forms of "expression cloning" have been extensively used for cloning receptors for protein ligands; see, for example, H. Simonsen and H.F. Lodish, Trends in Pharmacological Sciences, Volume 15, pages 437-441 (1994), and Tartaglia et al., Cell, Volume 83, pages 1263-1271 (1995). Such α2 receptor cloning experiments are described in greater detail in the "Detailed Description of the Invention" below. Isolation of the α2 receptor(s) would be very useful in terms of being able to identify or develop novel agonists and antagonists of the α2 signaling pathway. Such agonists and antagonists could include soluble α2 receptor(s), anti-α2-receptor antibodies, small molecules or antisense oligonucleotides, any of which can be used for potentially treating one or more of the diseases or disorders mentioned above.

Brief Description of the Figures

FIGURE 1. This figure depicts in linear fashion the full coding region of human and mouse α2 polypeptides in accordance with this invention. For

both proteins, the predicted signal peptide region is underscored and the region containing the predicted signal peptide cleavage site is boxed. The two asparagine (N) residues in each polypeptide that are located within classic NxT glycosylation motifs, and which are very likely to be glycosylated, are shown in larger font.

FIGURE 2. This figure shows the high degree of homology (GAP analysis) between the human and mouse α2 polypeptides of SEQ ID NOS: 1 and 2.

FIGURE 3A-3B. This figure illustrates the relatedness between human glycoprotein hormone α polypeptide (prior art) and the α2 polypeptide of this invention. The mature form of α2 used for this comparison (SEQ ID NO: 8) most likely represents the authentic *in vivo* form of α2. Figure 3A comprises the GAP output showing the amino acid homology between the mature forms of α and α2. Figure 3B illustrates the functional significance of the homology between the known α polypeptide and the α2 polypeptide of this invention in schematic form (long arrows link critically important amino acids in α with their homologs in α2. The six cysteines that form the cystine-knot in α are numbered 1 through 6 and are shown in large font. The three intramolecular disulfide bonds which they form are indicated by dashed lines. The locations of the two classic cystine-knot motifs, namely 2-3 CxGxC and 5-6 CxC, are indicated. Black rectangles indicate glycosylated asparagine NxT-motifs. The NxT-motif glycosylated asparagine (N) that has been shown in the α polypeptide of human glycoprotein hormones to be important for signal transduction through the glycoprotein hormone receptors is shown boxed, with the "N" and the "T" shown in large font. The

locations of intron two and intron three for both genes is indicated by appropriately labeled arrows.

FIGURE 4. This figure depicts the linear sequence of the human α2 cDNA (SEQ ID NO: 21) which consists of the 5′ untranslated region, the α2 coding region and the 3′ untranslated region (the polyA tail sequence is not shown but its location is indicated). The α2 polypeptide start codon (ATG, or methionine) and stop codon (TAG) are shown in large font capitals. The putative polyA-signal (consensus sequence "aataaa") is underlined. The location (but not the sequence) of the three introns in the corresponding human genomic sequence is indicated.

FIGURE 5. This figure shows the likely disulfide bond cysteine (C) pairs of the five putative disulfide bonds in human α2. The ten cysteine residues are shown in large font and the disulfide bonds are drawn as solid lines. The three disulfide bonds (C34-C96, C38-C98, C65-C101) that form the cystine-knot are drawn above the amino acid sequence and the two additional disulfide bonds (C8-C66, C25-C80) are drawn below the amino acid sequence.

FIGURE 6. This figure depicts the amino acid sequence of a fusion protein formed between the mouse α2 putative signal peptide region and human CD2 lacking its native signal peptide. This fusion protein (described in the text) was used to determine whether the putative signal peptide of α2 could function as a signal peptide in cultured 293T cells. The sequence from mouse α2 is shown in large font, with the predicted signal peptide underlined. The amino acids linking the mouse α2 sequence to the truncated human CD2 sequence are shown in italics. The truncated human CD2 sequence is boxed.

FIGURE 7A-7C. This figure shows the fluorescence of adherent 293T cells stained with an FITC-labeled anti-human-CD2 monoclonal antibody. Figure 7A, negative control: cells transfected with a plasmid containing a histone gene. Figure 7B, positive control: cells transfected with a plasmid containing the human CD2 gene. Figure 7C, cells transfected with a plasmid coding for the mouse-α2-human-CD2 fusion protein shown in Figure 6.

Detailed Description of the Invention

The section headings herein are for organizational purposes only and are not to be interpreted as limiting the subject matter described therein. All references cited in this application are expressly incorporated herein by reference.

Definitions

As used herein, the terms "polypeptide" and "protein" are meant to be interchangeable.

Related alpha 2 nucleic acid molecules include those molecules which comprise nucleotide sequences that hybridize under moderately or highly stringent conditions as defined herein with the complements of any of the above nucleic acid molecules. In preferred embodiments, the related nucleic acid molecules comprise sequences which hybridize under moderately or highly stringent conditions with the sequence as shown in SEQ ID NO: 21, or with a molecule encoding a polypeptide, which polypeptide comprises the sequence as shown in SEQ ID NO: 1, 2, and 5-12, or with a nucleic acid fragment as defined above, or with a nucleic acid fragment

encoding a polypeptide as defined above. It is also understood that related nucleic acid molecules include allelic or splice variants of any of the above nucleic acids, and include sequences which are complementary to any of the above nucleotide sequences.

The term "isolated nucleic acid molecule" refers to a nucleic acid molecule of the invention that is free from at least one contaminating nucleic acid molecule with which it is naturally associated, and preferably substantially free from any other contaminating mammalian nucleic acid molecules which would interfere with its use in protein production, or its therapeutic or diagnostic use, or any other uses which are described herein.

The term "allelic variant" refers to one of several possible naturally occurring alternate forms of a gene occupying a given locus on a chromosome of an organism or a population of organisms.

The term "splice variant" refers to a nucleic acid molecule, usually RNA, which is generated by alternative processing of intron sequences in an RNA transcript.

The term "expression vector" refers to a vector which is suitable for propagation in a host cell and contains nucleic acid sequences which direct and/or control the expression of inserted heterologous nucleic acid sequences. Expression includes, but is not limited to, processes such as transcription, translation, and RNA splicing, if introns are present.

The term "highly stringent conditions" refers to those conditions that (1) employ low ionic strength reagents and high temperature for washing,

for example, 0.015 M NaCl/0.0015 M sodium citrate/0.1% NaDodSO4 (SDS) at 50°C, or (2) employ during hybridization a denaturing agent such as formamide, for example, 50% (vol/vol) formamide with 0.1% bovine serum albumin/0.1%. Alternatively, an example includes use of Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 may be used with 750 mM NaCl, 75 mM sodium citrate at 42°C. Another example is the use of 50% formamide, 5 X SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 X Denhardt's solution, sonicated salmon sperm DNA (50 µg/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 X SSC and 0.1% SDS.

The term "moderately stringent conditions" refers to conditions which generally include the use of a washing solution and hybridization conditions (e.g., temperature, ionic strength, and percent SDS) less stringent than described above. For example, moderately stringent conditions are conditions such as overnight incubation at 37°C in a solution comprising 20% formamide, 5X SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5X Denhardt's solution, 10% dextran sulfate, and 20 micrograms/ml denatured sheared salmon sperm DNA, followed by washing in 1X SSC at about 37-50°C. The skilled artisan will recognize how to adjust the temperature, ionic strength etc. as necessary to accommodate factors such as probe length and the like.

In certain preferred embodiments, where oligonucleotide probes are used to screen cDNA or genomic libraries, high stringency conditions are used which depend upon the melting temperature (T_m) of oligonucleotide probes to target sequences. The T_m may be estimated using the formula shown below; for

reference, see Bolton et al. Proceedings of the National Academy of Sciences USA, Volume 48, page 1390 (1962).

$$T_m=81.5-16.6(\log[Na+])+0.41(\%G+C)-(600/N)$$

wherein [Na+] is the sodium ion concentration in the hybridization (or washing) solution;

% G+C is guanine and cytosine content in the oligonucleotide probe; and

N is the probe length in nucleotides.

An example of a high stringency solution is 6X SSC and 0.05% sodium pyrophosphate at a temperature of 35°C-63°C, depending on the length of the oligonucleotide probe. For example, according to certain embodiments, 14 base pair probes are washed at 35-40°C, 17 base pair probes are washed at 45-50°C, 20 base pair probes are washed at 52-57°C, and 23 base pair probes are washed at 57-63°C. The temperature can be increased 2-3°C where the background non-specific binding appears high. A second high stringency solution utilizes tetramethylammonium chloride (TMAC) for washing oligonucleotide probes. One stringent washing solution is 3 M TMAC, 50 mM Tris-HCl, pH 8.0, and 0.2 percent SDS. The washing temperature using this solution is a function of the length of the probe. For example, 14 base pair probes are washed at 35-40°C, 17 base pair probes are washed at about 45-50°C, 20 base pair probes are washed at 52-57°C, and 23 base pair probes are washed at 57-63°C.

The term "alpha 2 polypeptide fragment" refers to a peptide or polypeptide that comprises less than the full length amino acid sequence of an alpha 2 polypeptide as set forth in SEQ ID NO: 1, 2 and 5-12.

Such a fragment, which preferably comprises fifty or more amino acid residues, may arise, for example, from a truncation at the amino terminus, a truncation at the carboxy terminus, and/or an internal deletion of a residue(s) from the amino acid sequence. Alpha 2 fragments may result from alternative RNA splicing or from *in vivo* protease activity.

The term "alpha 2 polypeptide variants" refers to alpha 2 polypeptides comprising amino acid sequences which contain one or more amino acid sequence substitutions, deletions and/or additions as compared to the alpha 2 polypeptide amino acid sequence set forth in SEQ ID NO: 1, 2 and 5-12. Variants may be naturally occurring or artificially constructed. Such alpha 2 polypeptide variants may be prepared from the corresponding nucleic acid molecules encoding said variants, which have a DNA sequence that varies accordingly from the DNA sequences for wild type alpha 2 polypeptides as set forth in SEQ ID NOS: 1, 2 and 5-12.

The term "alpha 2 fusion polypeptide" refers to a fusion of alpha 2 polypeptide, fragment, variant and/or derivative thereof, with a heterologous peptide or polypeptide.

The term "alpha 2 polypeptide derivatives" refers to alpha 2 polypeptides, variants, or fragments thereof, that have been chemically modified, as for example, by covalent attachment of one or more polymers, including, but limited to, water soluble polymers, N-linked or O-linked carbohydrates, sugars, phosphates, and/or other such molecules. The derivatives are modified in a manner that is different from naturally occurring alpha 2, either in the type or the location of the molecules attached to the

polypeptide. Derivatives further includes deletion of one or more chemical groups naturally attached to the alpha 2 polypeptide.

In addition, an alpha 2 polypeptide may be active as an immunogen, that is, the polypeptide contains at least one epitope to which antibodies may be raised.

"Naturally occurring" when used in connection with biological materials such as nucleic acid molecules, polypeptides, host cells, and the like, refers to that which are found in nature and not manipulated by a human being.

The term "isolated polypeptide" refers to a polypeptide of the invention that is free from at least one contaminating polypeptide that is found in its natural environment, and preferably substantially free from any other contaminating mammalian polypeptides which would interfere with its therapeutic, diagnostic or other described use. This does not mean free from other proteins with which is normally functionally associated, such as heterodimerization with some other subunit.

The term "ortholog" refers to a polypeptide that corresponds to a polypeptide identified from a different species. For example, mouse and human alpha 2 polypeptides are considered orthologs.

The term "mature alpha 2 polypeptide" refers to a polypeptide lacking a leader sequence and may also include other modifications of a polypeptide such as proteolytic processing of the amino terminus (with or without a leader sequence) and/or the carboxy terminus, cleavage of a smaller polypeptide from a

larger precursor, N-linked and/or O-linked glycosylation, and the like.

The terms "effective amount" and "therapeutically effective amount" refer to the amount of an alpha 2 polypeptide which is useful or necessary to support an observable level of one or more biological activities of the alpha 2 polypeptides referred to above.

Relatedness of Nucleic Acid Molecules and/or Polypeptides

The term "identity", as is known in the art, refers to a relationship between the sequences of two or more polypeptide molecules or two or more nucleic acid molecules, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or nucleic acid molecule sequences, as the case may be, as determined by the match between strings of nucleotide or amino acid sequences. "Identity" measures the percent of identical matches between two or more sequences with gap alignments addressed by a particular mathematical model or computer programs (i.e., "algorithms").

The term "similarity" is a related concept, but in contrast to "identity", refers to a measure of similarity which includes both identical matches and conservative substitution matches. Since conservative substitutions apply to polypeptides and not nucleic acid molecules, similarity only deals with polypeptide sequence comparisons. If two polypeptide sequences have, for example, 10/20 identical amino acids, and the remainder are all non-conservative substitutions, then the percent identity and similarity would both be

50%. If, in the same example, there are 5 more positions where there are conservative substitutions, then the percent identity remains 50%, but the per cent similarity would be 75% (15/20). Therefore, in cases where there are conservative substitutions, the degree of similarity between two polypeptide sequences will be higher than the percent identity between those two sequences.

The term "conservative amino acid substitution" refers to a substitution of a native amino acid residue with a non-native residue such that there is little or no effect on the polarity or charge of the amino acid residue at that position. For example, a conservative substitution results from the replacement of a non-polar residue in a polypeptide with any other non-polar residue. Furthermore, any native residue in the polypeptide may also be substituted with alanine, as has been previously described for "alanine scanning mutagenesis". General rules for conservative amino acid substitutions are set forth in Table I, below.

Table I

Conservative Amino Acid Substitutions

Original Residues	Exemplary Substitutions	Preferred Substitutions
Ala	Val,Leu,Ile	Val
Arg	Lys,Gln,Asn	Lys
Asn	Gln,His,Lys,Arg	Gln
Asp	Glu	Glu
Cys	Ser	Ser
Gln	Asn	Asn
Glu	Asp	Asp
Gly	Pro,Ala	Ala

His	Asn,Gln,Lys,Arg	Arg
Ile	Leu,Val,Met,Ala, Phe,Norleucine	Leu
Leu	Norleucine,Ile, Val,Met,Ala,Phe	Ile
Lys	Arg,Gln,Asn	Arg
Met	Leu,Phe,Ile	Leu
Phe	Leu,Val,Ile,Ala, Tyr	Leu
Pro	Ala	Ala
Ser	Thr	Thr
Thr	Ser	Ser
Trp	Tyr,Phe	Tyr
Tyr	Trp,Phe,Thr,Ser	Phe
Val	Ile,Met,Leu,Phe, Ala,Norleucine	Leu

Conservative amino acid substitutions also encompass non-naturally occurring amino acid residues which are typically incorporated by chemical peptide synthesis rather than by synthesis in biological systems. These include peptidomimetics, and other reversed or inverted forms of amino acid moieties.

Conservative modifications to the amino acid sequence (and the corresponding modifications to the encoding nucleotides) are expected to produce alpha 2 having functional and chemical characteristics similar to those of naturally occurring alpha 2. In contrast, substantial modifications in the functional and/or chemical characteristics of alpha 2 may be accomplished by selecting substitutions that differ significantly in their effect on maintaining (a) the structure of the molecular backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the

molecule at the target site, or (c) the bulk of the side chain. Naturally occurring residues may be divided into groups based on common side chain properties as follows:

1) hydrophobic: norleucine, Met, Ala, Val, Leu, Ile;
2) neutral hydrophilic: Cys, Ser, Thr;
3) acidic: Asp, Glu;
4) basic: Asn, Gln, His, Lys, Arg;
5) residues that influence chain orientation: Gly, Pro; and
6) aromatic: Trp, Tyr, Phe.

Non-conservative substitutions may involve the exchange of a member of one of these classes for a member from another class. Such substituted residues may be introduced into regions of the human alpha 2 molecule that are homologous with non-human alpha 2, or into the non-homologous regions of the molecule.

Identity and similarity of related nucleic acid molecules and polypeptides can be readily calculated by known methods, including but not limited to those described in Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part 1, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M. Stockton Press, New York, 1991; and Carillo et al., SIAM J. Applied Math., 48, 1073 (1988).

Preferred methods to determine identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package, including GAP (Devereux et al., Nucleic Acids Research 12, 387 (1984); Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, and FASTA (Atschul et al., J. Mol. Biol. 215, 403-410 (1990)). The BLAST X program is publicly available from the National Center for Biotechnology Information (NCBI) and other sources, such as *BLAST Manual,* Altschul, S., et al. NCB NLM NIH Bethesda, MD 20894; Altschul et al., above. The well known Smith Waterman algorithm may also be used to determine identity.

Certain alignment schemes for aligning two amino acid sequences may result in matching of only a short region of the two sequences, and this small aligned region may have very high sequence identity even though there is no significant relationship between the two full length sequences. Accordingly, in a preferred embodiment, the selected alignment method (GAP program) will result in an alignment that spans at least fifty contiguous amino acids of the claimed polypeptide.

By way of example, using the computer algorithm known as GAP (Genetics Computer Group, University of Wisconsin, Madison, WI), two polypeptides for which the percent sequence identity is to be determined are aligned for optimal matching of their respective amino acids (the "matched span", as determined by the algorithm). A gap opening

penalty (which is calculated as 3 X the average diagonal; the "average diagonal" is the average of the diagonal of the comparison matrix being used; the "diagonal" is the score or number assigned to each perfect amino acid match by the particular comparison matrix) and a gap extension penalty (which is usually 1/10 times the gap opening penalty), as well as a comparison matrix such as PAM 250 or BLOSUM 62 are used in conjunction with the algorithm. A standard comparison matrix (see Dayhoff et al. in: Atlas of Protein Sequence and Structure, vol. 5, supp.3 (1978) for the PAM250 comparison matrix; see Henikoff et al., Proc. Natl. Acad. Sci USA, 89, 10915-10919 (1992) for the BLOSUM 62 comparison matrix) is also used by the algorithm.

Preferred parameters for polypeptide sequence comparison include the following:

Algorithm: Needleman and Wunsch, J. Mol. Biol. 48, 443-453 (1970);
Comparison matrix: BLOSUM 62 from Henikoff et al., Proc. Natl. Acad. Sci. USA 89, 10915-10919 (1992);
Gap Penalty: 12;
Gap Length Penalty: 4;
Threshold of Similarity: 0

The GAP program is useful with the above parameters. The aforementioned parameters are the default parameters for polypeptide comparisons (along with no penalty for end gaps) using the GAP algorithm.

Preferred parameters for nucleic acid molecule sequence comparison include the following:

Algorithm: Needleman et al. J. Mol Biol. 48, 443-453 (1970);

Comparison matrix: matches = +10, mismatch = 0
Gap Penalty: 50;
Gap Length Penalty: 3.

The GAP program is also useful with the above parameters. The aforementioned parameters are the default parameters for nucleic acid molecule comparisons.

Other exemplary algorithms, gap opening penalties, gap extension penalties, comparison matrices, thresholds of similarity, etc., may be used by those of skill in the art, including those set forth in the Program Manual, Wisconsin Package, Version 9, September, 1997. The particular choices to be made will depend on the specific comparison to be made, such as DNA to DNA, protein to protein, protein to DNA; and additionally, whether the comparison is between given pairs of sequences (in which case GAP or BestFit are generally preferred) or between one sequence and a large database of sequences (in which case FASTA or BLASTA are preferred).

As referred to in the "Summary of Invention", above, there are very significant structural and functional similarities between the α2 protein of this invention and the human glycoprotein hormone α polypeptide. The mature form of the human glycoprotein hormone α polypeptide contains ten cysteines, six of which form the cystine-knot (Figure 3B). The cystine-knot is the critical structural feature that determines the overall three-dimensional structure of the glycoprotein hormone alpha polypeptide. Mature form human α2 also contains ten cysteines. Six of these are extremely likely to form a cystine-knot based upon their relative location within the α2 polypeptide sequence and the presence of

the two classic 2-3 CxGxC and 5-6 CxC cystine-knot motifs (Figure 3B). Mature form human glycoprotein hormone α polypeptide contains two NxT motif glycosylated asparagines (N), located at positions 52 and 78 (Figure 3B). Glycosylation at asparagine 52 has been shown to be important for signal transduction through the glycoprotein hormone receptors; see Matzuk and Boime, *Biology of Reproduction*, vol. 40, pp. 48-53 (1989); and in the crystal structure of CG this glycosylation site is within the proposed receptor-binding domain. Glycosylation at asparagine 78 is thought to play a role in determining the half-life of circulating hormone, consistent with its location on the solvent-exposed face of the hormone; Thotakura and Blithe, *Glycobiology*, vol. 5, pp. 3-10 (1995), Szkudlinski et al., *Endocrinology*, vol. 136, pp. 3325-3330 (1995), and Lapthorn et al., *Nature*, vol. 369, pp. 455-461 (1994). Mature form human α2 also contains two NxT motif asparagines (N), which are located at positions 14 and 58 in SEQ ID NO: 8. Asparagine 58 is positionally homologous to asparagine 52 of the human α polypeptide, implying conservation of an important part of the signal transduction mechanism. Asparagine 14 of human α2 is likely to be the functional equivalent of asparagine 78 in human α, due to the fact that asparagine 14 is in a region of human α2 that, in human α, is predicted by the CG crystal structure to be solvent-exposed.

The gene encoding the human α polypeptide contains four exons and three introns; see Fiddes and Goodman, *Journal of Molecular and Applied Genetics*. vol. 1, pp. 3-18 (1981). It has been determined that the gene encoding human α2 also contains four exons and three introns (See Figure 4). For both genes, the locations of the introns with respect to the major

structural/functional regions are fully conserved as follows:

a) Intron 1 is located just 5' of the start ATG (seven nucleotides 5' for α; zero nucleotides 5' for α2)

b) Intron 2 is located between the 5' end of the mature protein and the CxGxC cystine-knot motif. Intron 2 in both proteins splits a codon 1:2.

c) Intron 3 is located between the signal transduction NxT-motif asparagine and the CxC cystine-knot motif. Intron 3 in both proteins does not split a codon but lies between codons.

Mature human α polypeptide contains ten cysteine residues which form five intramolecular disulfide bonds. Mature human α2 also contains ten cysteine residues, which are likely to similarly form five disulfide bonds. Three-dimensional computer modeling of human α2 based on the crystal structure of human α supports the prediction that, most likely, the disulfide bond cysteine pairing for the five putative disulfide bonds in human α2 is as follows: C8-C66, C25-C80, C34-C96, C38-C98 and C65-C101 of SEQ ID NO: 8 (shown in Figure 5). A less likely, but still possible, disulfide bond cysteine pairing is: C8-C80, C25-C66, C34-C96, C38-C98 and C65-C101.

Another important property relating to the structure and function of α2 is that of glycosylation. Each of the known glycoprotein hormone polypeptides is glycosylated at one or more amino acid residues. Glycoprotein hormone α polypeptide is glycosylated at

two asparagine (N) residues, both of which are located within classic NxT glycosylation motifs. The predicted amino acid sequence of α2 contains two asparagines (N) in NxT motifs. By analogy to α, both of these NxT motif asparagine residues are very likely to be glycosylated. For the known glycoprotein hormones, glycosylation is known to greatly influence circulating half life (and thus *in vivo* biological activity), the extent of signal transduction in cultured mammalian cells expressing the appropriate glycoprotein hormone receptors, and the levels of hormone produced from various cell lines. There is a distinct possibility that the specific nature of the glycosylation on the two NxT motif asparagine residues in α2 will impact the same biological parameters influenced by glycosylation of the known glycoprotein hormones.

Nucleic Acid Molecules

Recombinant DNA methods used herein are generally those set forth in Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989) and/or Ausubel et al., eds., Current Protocols in Molecular Biology, Green Publishers Inc. and Wiley and Sons, NY (1994).

The invention provides for nucleic acid molecules as described herein and methods for obtaining the molecules. A gene or cDNA encoding an alpha 2 polypeptide or fragment thereof may be obtained by hybridization screening of a genomic or cDNA library, or by PCR amplification. Probes or primers useful for screening a library by hybridization can be generated based on sequence information for other known genes or gene fragments

from the same or a related family of genes, such as, for example, conserved motifs. In addition, where a gene encoding alpha 2 polypeptide has been identified from one species, all or a portion of that gene may be used as a probe to identify corresponding genes from other species (orthologs) or related genes from the same species (homologs). The probes or primers may be used to screen cDNA libraries from various tissue sources believed to express the alpha 2 gene. In addition, part or all of a nucleic acid molecule having the sequence as set forth in SEQ ID NO: 21 may be used to screen a genomic library to identify and isolate a gene encoding alpha 2. Typically, conditions of moderate or high stringency will be employed for screening to minimize the number of false positives obtained from the screen.

Nucleic acid molecules encoding alpha 2 polypeptides may also be identified by expression cloning which employs detection of positive clones based upon a property of the expressed protein. Typically, nucleic acid libraries are screened by binding of an antibody or other binding partner (e.g., receptor or ligand) to cloned proteins which are expressed and displayed on the host cell surface. The antibody or binding partner is modified with a detectable label to identify those cells expressing the desired clone.

Another means of preparing a nucleic acid molecule encoding an alpha 2 polypeptide or fragment thereof is chemical synthesis using methods well known to the skilled artisan such as those described by Engels et al., Angew. Chem. Intl. Ed., 28, 716-734 (1989). These methods include, *inter alia*, the phosphotriester, phosphoramidite, and H-phosphonate methods for nucleic acid synthesis. A preferred

method for such chemical synthesis is polymer-supported synthesis using standard phosphoramidite chemistry. Typically, the DNA encoding the alpha 2 polypeptide will be several hundred nucleotides in length. Nucleic acids larger than about 100 nucleotides can be synthesized as several fragments using these methods. The fragments can then be ligated together to form the full length alpha 2 polypeptide. Usually, the DNA fragment encoding the amino terminus of the polypeptide will have an ATG, which encodes a methionine residue. This methionine may or may not be present on the mature form of the alpha 2 polypeptide, depending on whether the polypeptide produced in the host cell is designed to be secreted from that cell.

In some cases, it may be desirable to prepare nucleic acid molecules encoding alpha 2 polypeptide variants. Nucleic acid molecules encoding variants may be produced using site directed mutagenesis, PCR amplification, or other appropriate methods, where the primer(s) have the desired point mutations (see Sambrook et al., above, and Ausubel et al., above, for descriptions of mutagenesis techniques). Chemical synthesis using methods described by Engels et al., above, may also be used to prepare such variants. Other methods known to the skilled artisan may be used as well.

In certain embodiments, nucleic acid variants contain codons which have been altered for optimal expression of an alpha 2 polypeptide in a given host cell. Particular codon alterations will depend upon the alpha 2 polypeptide(s) and host cell(s) selected for expression. Such "codon optimization" can be carried out by a variety of methods, for example, by selecting codons which are

preferred for use in highly expressed genes in a given host cell. Computer algorithms which incorporate codon frequency tables such as "Ecohigh. Cod" for codon preference of highly expressed bacterial genes may be used and are provided by the University of Wisconsin Package Version 9.0, Genetics Computer Group, Madison, WI. Other useful codon frequency tables include "Celegans_high.cod", "Celegans_low.cod", "Drosophila_high.cod", "Human_high.cod", "Maize_high.cod", and "Yeast_high.cod".

In other embodiments, nucleic acid molecules encode alpha 2 variants with conservative amino acid substitutions as defined above, alpha 2 variants comprising an addition and/or a deletion of one or more N-linked or O-linked glycosylation sites, alpha 2 variants having deletions and/or substitutions of one or more cysteine residues, or alpha 2 polypeptide fragments as described above. In addition, nucleic acid molecules may encode any combination of alpha 2 variants, fragments, and fusion polypeptides described herein.

Vectors and Host Cells

For purposes of recombinant manufacture of alpha 2 in accordance with this invention, a nucleic acid molecule encoding an alpha 2 polypeptide is inserted into an appropriate expression vector using standard ligation techniques. The vector is typically selected to be functional in the particular host cell employed (i.e., the vector is compatible with the host cell machinery such that amplification of the gene and/or expression of the gene can occur). A nucleic acid molecule encoding an alpha 2 polypeptide may be amplified/expressed in prokaryotic, yeast, insect

(baculovirus systems) and/or eukaryotic host cells. Selection of the host cell will depend in part on whether an alpha 2 polypeptide is to be post-translationally modified (e.g., glycosylated and/or phosphorylated). If so, yeast, insect or mammalian host cells are preferable. For a review of expression vectors, see Meth. Enz. v. 185, D.V. Goeddel, ed., Academic Press Inc., San Diego, CA (1990).

Typically, expression vectors used in any of the host cells will contain sequences for plasmid maintenance and for cloning and expression of exogenous nucleotide sequences. Such sequences, collectively referred to as "flanking sequences", in certain embodiments will typically include one or more of the following nucleotides: a promoter, one or more enhancer sequences, an origin of replication, a transcriptional termination sequence, a complete intron sequence containing donor and acceptor splice sites, a leader sequence for secretion, a ribosome binding site, a polyadenylation sequence, a polylinker region for inserting the nucleic acid encoding the polypeptide to be expressed, and a selectable marker element. Each of these sequences is discussed below.

Optionally, the vector may contain a "tag" sequence, i.e., an oligonucleotide molecule located at the 5' or 3' end of the alpha 2 polypeptide coding sequence; the oligonucleotide molecule encodes polyHis (such as hexaHis), or some other "tag" such as FLAG, HA (hemaglutinin Influenza virus) or *myc*, for which commercially available antibodies exist. This tag is typically fused to the polypeptide upon expression of the latter, and can serve as a means for affinity purification of the alpha 2 polypeptide from the host cell. Affinity purification can be accomplished, for example, by column chromatography using antibodies

against the tag as an affinity matrix. Optionally, the tag can subsequently be removed from the purified alpha 2 polypeptide by various means, such as using certain peptidases for cleavage.

Flanking sequences may be homologous (i.e., from the same species and/or strain as the host cell), heterologous (i.e., from a species other than the host cell species or strain), hybrid (i.e., a combination of flanking sequences from more than one source), or synthetic, or native sequences which normally function to regulate alpha 2 expression. As such, the source of flanking sequences may be any prokaryotic or eukaryotic organism, any vertebrate or invertebrate organism, or any plant, provided that the flanking sequences is functional in, and can be activated by, the host cell machinery.

The flanking sequences useful in the vectors of this invention may be obtained by any of several methods well known in the art. Typically, flanking sequences useful herein other than the alpha 2 gene flanking sequences will have been previously identified by mapping and/or by restriction endonuclease digestion and can thus be isolated from the proper tissue source using the appropriate restriction endonucleases. In some cases, the full nucleotide sequence of one or more flanking sequence may be known. Here, the flanking sequence may be synthesized using the methods described above for nucleic acid synthesis or cloning.

Where all or only a portion of the flanking sequence is known, it may be obtained using PCR and/or by screening a genomic library with suitable oligonucleotide and/or flanking sequence fragments from the same or another species.

Where the flanking sequence is not known, a fragment of DNA containing a flanking sequence may be isolated from a larger piece of DNA that may contain, for example, a coding sequence or even another gene or genes. Isolation may be accomplished by restriction endonuclease digestion to produce the proper DNA fragment followed by isolation using agarose gel purification, Qiagen® column chromatography, or other methods known to the skilled artisan. Selection of suitable enzymes to accomplish this purpose will be readily apparent to one of ordinary skill in the art.

An origin of replication is typically a part of prokaryotic expression vectors purchased commercially, and it aids in the amplification of the vector in a host cell. Amplification of the vector to a certain copy number can, in some cases, be important for optimal expression of the alpha 2 polypeptide. If the vector of choice does not contain an origin of replication site, one may be chemically synthesized based on a known sequence and ligated into the vector.

The origin of replication from the plasmid pBR322 (Product No. 303-3s, New England Biolabs, Beverly, MA) is suitable for most Gram-negative bacteria and various origins (e.g., SV40, polyoma, adenovirus, vesicular stomatitus virus, or papillomaviruses such as HPV or BPV) are useful for cloning vectors in mammalian cells. Generally, the origin of replication component is not needed for mammalian expression vectors (for example, the SV40 origin is often used only because it contains the early promoter).

A transcription termination sequence is typically located 3' of the end of a polypeptide

coding regions and serves to terminate transcription. Usually, a transcription termination sequence in prokaryotic cells is a G-C rich fragment followed by a poly T sequence. While the sequence is easily cloned from a library or even purchased commercially as part of a vector, it can also be readily synthesized using methods for nucleic acid synthesis such as those described above.

A selectable marker gene element encodes a protein necessary for the survival and growth of a host cell grown in a selective culture medium. Typical selection marker genes encode proteins that (a) confer resistance to antibiotics or other toxins, e.g., ampicillin, tetracycline, or kanamycin for prokaryotic host cells, (b) complement auxotrophic deficiencies of the cell, or (c) supply critical nutrients not available from complex media. Preferred selectable markers are the kanamycin resistance gene, the ampicillin resistance gene and the tetracycline resistance gene. A neomycin resistance gene may also be used for selection in prokaryotic and eukaryotic host cells.

Other selection genes may be used to amplify the gene to be expressed. Amplification is the process wherein genes which are in greater demand for the production of a protein critical for growth are reiterated in tandem within the chromosomes of successive generations of recombinant cells. Examples of suitable selectable markers for mammalian cells include dihydrofolate reductase (DHFR) and thymidine kinase. The mammalian cell transformants are placed under selection pressure which only the transformants are uniquely adapted to survive by virtue of the marker present in the vector. Selection pressure is imposed by culturing the transformed cells under

conditions in which the concentration of selection agent in the medium is successively changed, thereby leading to amplification of both the selection gene and the DNA that encodes alpha 2. As a result, increased quantities of alpha 2 are synthesized from the amplified DNA.

A ribosome binding site is usually necessary for translation initiation of mRNA and is characterized by a Shine-Dalgarno sequence (prokaryotes) or a Kozak sequence (eukaryotes). The element is typically located 3' to the promoter and 5' to the coding sequence of the alpha 2 polypeptide to be expressed. The Shine-Dalgarno sequence is varied but is typically a polypurine (i.e., having a high A-G content). Many Shine-Dalgarno sequences have been identified, each of which can be readily synthesized using methods set forth above and used in a prokaryotic vector.

A leader, or signal, sequence may be used to direct an alpha 2 polypeptide out of the host cell. Typically, the signal sequence is positioned in the coding region of the alpha 2 nucleic acid molecule, or directly at the 5' end of the alpha 2 polypeptide coding region. Many signal sequences have been identified, and any of them that are functional in the selected host cell may be used in conjunction with the alpha 2 gene or cDNA. Therefore, a signal sequence may be homologous (naturally occurring) or heterologous to the alpha 2 gene or cDNA, Additionally, a signal sequence may be chemically synthesized using methods set forth above. In most cases, secretion of an alpha 2 polypeptide from the host cell via the presence of a signal peptide will result in the removal of the signal peptide from the alpha 2 polypeptide. The signal sequence may be a

component of the vector, or it may be a part of alpha 2 DNA that is inserted into the vector.

Included within the scope of this invention is the native alpha 2 signal sequence joined to an alpha 2 coding region and a heterologous signal sequence joined to an alpha 2 coding region. The heterologous signal sequence selected should be one that is recognized and processed, i.e., cleaved by a signal peptidase, by the host cell. For prokaryotic host cells that do not recognize and process the native alpha 2 signal sequence, the signal sequence is substituted by a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, or heat-stable enterotoxin II leaders. In mammalian cell expression the native signal sequence is satisfactory, although other mammalian signal sequences may be suitable. For yeast secretion, the native alpha 2 signal sequence may be substituted by the yeast invertase, alpha factor, or acid phosphatase leaders.

In some cases, such as where glycosylation is desired in a eukaryotic host cell expression system, one may manipulate the various presequences to improve glycosylation or yield. For example, one may alter the peptidase cleavage site of a particular signal peptide, or add prosequences, which also may affect glycosylation. The final protein product may have, in the -1 position (relative to the first amino acid of the mature protein) one or more additional amino acids incident to expression, which may not have been totally removed. For example, the final protein product may have one or two amino acid found in the peptidase cleavage site, attached to the N-terminus. Alternatively, use of some enzyme cleavage sites may result in a slightly truncated form of the desired

alpha 2 polypeptide, if the enzyme cuts at such area within the mature polypeptide.

In many cases, transcription of a nucleic acid molecule is increased by the presence of one or more introns in the vector; this is particularly true where a polypeptide is produced in eukaryotic host cells, especially mammalian host cells. The introns used may be naturally occurring within the alpha 2 gene, especially where the gene used is a full length genomic sequence or a fragment thereof. Where the intron is not naturally occurring within the gene (as for most cDNAs), the intron(s) may be obtained from another source. The position of the intron with respect to flanking sequences and the alpha 2 gene is generally important, as the intron must be transcribed to be effective. Thus, when an alpha 2 cDNA molecule is being expressed, the preferred position for the intron is 3' to the transcription start site, and 5' to the polyA transcription termination sequence. Preferably, the intron or introns will be located on one side or the other (i.e., 5' or 3') of the cDNA such that it does not interrupt the coding sequence. Any intron from any source, including any viral, prokaryotic and eukaryotic (plant or animal) organisms, may be used to practice this invention, provided that it is compatible with the host cell into which it is inserted. Also included herein are synthetic introns. Optionally, more than one intron may be used in the vector.

The expression and cloning vectors of the present invention will typically contain a promoter that is recognized by the host organism and operably linked to the molecule encoding the alpha 2 protein. Promoters are untranslated sequences located upstream (i.e., 5') to the start codon of a structural gene

(generally within about 100 to 1000 base pairs) that control the transcription and translation of the structural gene. Promoters are conventionally grouped into one of two classes, inducible promoters and constitutive promoters. Inducible promoters initiate increased levels of transcription from DNA under their control in response to some change in culture conditions, such as the presence or absence of a nutrient or a change in temperature. A large number of promoters, recognized by a variety of potential host cells, are well known. These promoters are operably linked to the DNA encoding alpha 2 by removing the promoter from the source DNA by restriction enzyme digestion and inserting the desired promoter sequence into the vector. The native alpha 2 promoter sequence may be used to direct amplification and/or expression of alpha 2 DNA. A heterologous promoter is preferred, however, if it permits greater transcription and higher yields of the expressed protein as compared to the native promoter, and if it is compatible with the host cell system that has been selected for use.

Promoters suitable for use with prokaryotic hosts include the beta-lactamase and lactose promoter systems; alkaline phosphatase, a tryptophan (trp) promoter system; and hybrid promoters such as the tac promoter. Other known bacterial promoters are also suitable. Their sequences have been published, thereby enabling one skilled in the art to ligate them to the desired DNA sequence(s), using linkers or adapters as needed to supply any required restriction sites.

Suitable promoters for use with yeast hosts are also well known in the art. Yeast enhancers are advantageously used with yeast promoters. Suitable

promoters for use with mammalian host cells are well known and include those obtained from the genomes of viruses such as polyoma virus, fowlpox virus, adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and most preferably Simian Virus 40 (SV40). Other suitable mammalian promoters include heterologous mammalian promoters, e.g., heat-shock promoters and the actin promoter.

Additional promoters which may be of interest in controlling alpha 2 expression include, but are not limited to: the SV40 early promoter region (Bernoist and Chambon, Nature, 290:304-310, 1981); the CMV promoter; the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto, et al., Cell, 22:787-797, 1980); the herpes thymidine kinase promoter (Wagner et al., Proc. Natl. Acad. Sci. U.S.A., 78:144-1445, 1981); the regulatory sequences of the metallothionine gene (Brinster et al., Nature, 296:39-42, 1982); prokaryotic expression vectors such as the beta-lactamase promoter (Villa-Kamaroff, et al., Proc. Natl. Acad. Sci. U.S.A., 75:3727-3731, 1978); or the tac promoter (DeBoer, et al., Proc. Natl. Acad. Sci. U.S.A., 80:21-25, 1983). Also of interest are the following animal transcriptional control regions which exhibit tissue specificity and have been utilized in transgenic animals: the elastase I gene control region which is active in pancreatic acinar cells (Swift et al., Cell, 38:639-646, 1984; Ornitz et al., Cold Spring Harbor Symp. Quant. Biol. 50:399-409, 1986; MacDonald, Hepatology, 7:425-515, 1987); the insulin gene control region which is active in pancreatic beta cells (Hanahan, Nature, 315:115-122, 1985); the immunoglobulin gene control region which is active in lymphoid cells (Grosschedl et al., Cell, 38:647-658,

1984; Adames et al., Nature, 318:533-538, 1985; Alexander et al., Mol. Cell. Biol., 7:1436-1444, 1987); the mouse mammary tumor virus control region which is active in testicular, breast, lymphoid and mast cells (Leder et al., Cell, 45:485-495, 1986), albumin gene control region which is active in liver (Pinkert et al., Genes and Devel., 1:268-276, 1987); the alphafetoprotein gene control region which is active in liver (Krumlauf et al., Mol. Cell. Biol., 5:1639-1648, 1985; Hammer et al., Science, 235:53-58, 1987); the alpha 1-antitrypsin gene control region which is active in the liver (Kelsey et al., Genes and Devel., 1:161-171, 1987); the beta-globin gene control region which is active in myeloid cells (Mogram et al., Nature, 315:338-340, 1985; Kollias et al., Cell, 46:89-94, 1986); the myelin basic protein gene control region which is active in oligodendrocyte cells in the brain (Readhead et al., Cell, 48:703-712, 1987); the myosin light chain-2 gene control region which is active in skeletal muscle (Sani, Nature, 314:283-286, 1985); and the gonadotropic releasing hormone gene control region which is active in the hypothalamus (Mason et al., Science, 234:1372-1378, 1986).

An enhancer sequence may be inserted into the vector to increase the transcription of a DNA encoding an alpha 2 protein of the present invention by higher eukaryotes. Enhancers are cis-acting elements of DNA, usually about 10-300 base pairs in length, that act on the promoter to increase its transcription. Enhancers are relatively orientation and position independent. They have been found 5' and 3' to the transcription unit. Several enhancer sequences available from mammalian genes are known (e.g., globin, elastase, albumin, alpha-feto-protein and insulin). Typically, however, an enhancer from a virus will be used. The SV40 enhancer, the

cytomegalovirus early promoter enhancer, the polyoma enhancer, and adenovirus enhancers are exemplary enhancing elements for the activation of eukaryotic promoters. While an enhancer may be spliced into the vector at a position 5' or 3' to alpha 2 DNA, it is typically located at a site 5' from the promoter.

Expression vectors of the invention may be constructed from a starting vectors such as a commercially available vector. Such vectors may or may not contain all of the desired flanking sequences. Where one or more of the flanking sequences set forth above are not already present in the vector to be used, they may be individually obtained and ligated into the vector. Methods used for obtaining each of the flanking sequences are well known to one skilled in the art.

Preferred vectors for practicing this invention are those which are compatible with bacterial, insect and mammalian host cells. Such vectors include, *inter alia*, pCRII, pCR3, and pcDNA3.1 (Invitrogen Company, San Diego, CA), pBSII (Stratagene Company, La Jolla, CA), pET15β (Novagen, Madison, WI), pGEX (Pharmacia Biotech, Piscataway, NJ), pEGFP-N2 (Clontech, Palo Alto, CA), pETL (BlueBacII; Invitrogen), pDSR-alpha (PCT Publication No. WO90/14363) and pFastBacDual (Gibco/BRL, Grand Island, NY).

Additional possible vectors include, but are not limited to, cosmids, plasmids or modified viruses, provided the vector system is compatible with the selected host cell. Such vectors include, but are not limited to, plasmids such as Bluescript® plasmid derivatives (a high copy number ColE1-based phagemid, Stratagene Cloning Systems Inc., La Jolla CA), PCR

cloning plasmids designed for cloning Taq-amplified PCR products (e.g., TOPO™ TA Cloning® Kit, PCR2.1® plasmid derivatives, Invitrogen, Carlsbad, CA), and mammalian, yeast or virus vectors such as a baculovirus expression system (pBacPAK plasmid derivatives, Clontech, Palo Alto, CA). The recombinant molecules can be introduced into host cells via transformation, transfection, infection, electroporation, or other known techniques.

After the vector has been constructed and a nucleic acid molecule encoding an alpha 2 polypeptide has been inserted into the proper site of the vector, the completed vector may be inserted into a suitable host cell for amplification and/or polypeptide expression.

Host cells may be prokaryotic host cells (such as *E. coli*) or eukaryotic host cells (such as a yeast cell, an insect cell, or a vertebrate cell). The host cell, when cultured under appropriate conditions, synthesizes an alpha 2 polypeptide which can subsequently be collected from the culture medium (if the host cell secretes it into the medium) or directly from the host cell producing it (if it is not secreted). Selection of an appropriate host cell will depend upon various factors, such as desired expression levels, polypeptide modifications that are desirable or necessary for activity, such as glycosylation or phosphorylation, and ease of folding into a biologically active molecule.

A number of suitable host cells are known in the art and many are available from the American Type Culture Collection (ATCC), Manassas, VA. Examples include mammalian cells, such as Chinese hamster ovary cells (CHO) (ATCC No. CCL61) CHO DHFR- cells (Urlaub

et al. Proc. Natl. Acad. Sci. USA 97, 4216-4220 (1980)), human embryonic kidney (HEK) 293 or 293T cells (ATCC No. CRL1573), or 3T3 cells (ATCC No. CCL92). The selection of suitable mammalian host cells and methods for transformation, culture, amplification, screening and product production and purification are known in the art. Other suitable mammalian cell lines are the monkey COS-1 (ATCC No. CRL1650) and COS-7 cell lines (ATCC No. CRL1651), and the CV-1 cell line (ATCC No. CCL70). Further exemplary mammalian host cells include primate cell lines and rodent cell lines, including transformed cell lines. Normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, as well as primary explants, are also suitable. Candidate cells may be genotypically deficient in the selection gene, or may contain a dominantly acting selection gene. Other suitable mammalian cell lines include but are not limited to, mouse neuroblastoma N2A cells, HeLa, mouse L-929 cells, 3T3 lines derived from Swiss, Balb-c or NIH mice, BHK or HaK hamster cell lines, which are available from the American Type Culture Collection, Manassas, VA). Each of these cell lines is known by and available to those skilled in the art of protein expression.

Similarly useful as host cells suitable for the present invention are bacterial cells. For example, the various strains of *E. coli* (e.g., HB101, (ATCC No. 33694) DH5α, DH10, and MC1061 (ATCC No. 53338)) are well-known as host cells in the field of biotechnology. Various strains of B. subtilis, Pseudomonas spp., other Bacillus spp., Streptomyces spp., and the like may also be employed in this method.

Many strains of yeast cells known to those skilled in the art are also available as host cells for expression of the polypeptides of the present invention. Preferred yeast cells include, for example, *Saccharomyces cerivisae*.

Additionally, where desired, insect cell systems may be utilized in the methods of the present invention. Such systems are described, for example, in Kitts et al. (Biotechniques, 14, 810-817 (1993)), Lucklow (Curr. Opin. Biotechnol., 4, 564-572 (1993) and Lucklow et al. (J. Virol., 67, 4566-4579 (1993)). Preferred insect cells are Sf-9 and Hi5 (Invitrogen, Carlsbad, CA).

Transformation or transfection of an expression vector for an alpha 2 polypeptide into a selected host cell may be accomplished by well known methods including methods such as calcium chloride, electroporation, microinjection, lipofection or the DEAE-dextran method. The method selected will in part be a function of the type of host cell to be used. These methods and other suitable methods are well known to the skilled artisan, and are set forth, for example, in Sambrook et al., above.

One may also use transgenic animals to express glycosylated alpha 2 polypeptides. For example, one may use a transgenic milk-producing animal (a cow or goat, for example) and obtain the present glycosylated polypeptide in the animal milk. One may also use plants to produce alpha 2 polypeptides, however, in general, the glycosylation occurring in plants is different from that produced in mammalian cells, and may result in a glycosylated product which is not suitable for human therapeutic use.

Polypeptide Production

Host cells comprising an alpha 2 expression vector (i.e., transformed or transfected) may be cultured using standard media well known to the skilled artisan. The media will usually contain all nutrients necessary for the growth and survival of the cells. Suitable media for culturing *E. coli* cells are for example, Luria Broth (LB) and/or Terrific Broth (TB). Suitable media for culturing eukaryotic cells are RPMI 1640, MEM, DMEM, all of which may be supplemented with serum and/or growth factors as required by the particular cell line being cultured. A suitable medium for insect cultures is Grace's medium supplemented with yeastolate, lactalbumin hydrolysate, and/or fetal calf serum as necessary.

Typically, an antibiotic or other compound useful for selective growth of transfected or transformed cells is added as a supplement to the media. The compound to be used will be dictated by the selectable marker element present on the plasmid with which the host cell was transformed. For example, where the selectable marker element is kanamycin resistance, the compound added to the culture medium will be kanamycin. Other compounds for selective growth include ampicillin, tetracycline and neomycin.

The amount of an alpha 2 polypeptide produced by a host cell can be evaluated using standard methods known in the art. Such methods include, without limitation, Western blot analysis, SDS-polyacrylamide gel electrophoresis, non-denaturing gel electrophoresis, HPLC separation,

immunoprecipitation, and/or activity assays such as DNA binding gel shift assays.

If an alpha 2 polypeptide has been designed to be secreted from the host cells, the majority of polypeptide may be found in the cell culture medium. If, however, the alpha 2 polypeptide is not secreted from the host cells, it will be present in the cytoplasm and/or the nucleus (for eukaryotic host cells) or in the cytosol (for gram negative bacteria host cells).

For an alpha 2 polypeptide situated in the host cell cytoplasm and/or nucleus, the host cells are typically first disrupted mechanically or with detergent to release the intra-cellular contents into a buffered solution. Alpha 2 polypeptide can then be isolated from this solution. If part of a heterodimer, alpha 2 can be isolated from this solution in heterodimeric form, as those skilled in the art will understand.

Purification of an alpha 2 polypeptide from solution, either as a monomer or dimer (homo or hetero), can be accomplished using a variety of techniques. If the polypeptide has been synthesized such that it contains a tag such as Hexahistidine (alpha 2 polypeptide/hexaHis) or other small peptide such as FLAG (Eastman Kodak Co., New Haven, CT) or *myc* (Invitrogen) at either its carboxyl or amino terminus, it may essentially be purified in a one-step process by passing the solution through an affinity column where the column matrix has a high affinity for the tag or for the polypeptide directly (i.e., a monoclonal antibody specifically recognizing alpha 2 polypeptide). For example, polyhistidine binds with great affinity and specificity to nickel, thus an

affinity column of nickel (such as the Qiagen® nickel columns) can be used for purification of alpha 2 polypeptide/polyHis. See, for example, Ausubel et al., eds., *Current Protocols in Molecular Biology*, Section 10.11.8, John Wiley & Sons, New York (1993).

Where an alpha 2 polypeptide is prepared without a tag attached, and no antibodies are available, other well known procedures for purification can be used. Such procedures include, without limitation, ion exchange chromatography, molecular sieve chromatography, HPLC, native gel electrophoresis in combination with gel elution, and preparative isoelectric focusing ("Isoprime" machine/technique, Hoefer Scientific). In some cases, two or more of these techniques may be combined to achieve increased purity.

If an alpha 2 polypeptide is produced intracellularly, the intracellular material (including inclusion bodies for gram-negative bacteria) can be extracted from the host cell using any standard technique known to the skilled artisan. For example, the host cells can be lysed to release the contents of the periplasm/cytoplasm by French press, homogenization, and/or sonication followed by centrifugation.

If an alpha 2 polypeptide has formed inclusion bodies in the cytosol, the inclusion bodies can often bind to the inner and/or outer cellular membranes and thus will be found primarily in the pellet material after centrifugation. The pellet material can then be treated at pH extremes or with chaotropic agent such as a detergent, guanidine, guanidine derivatives, urea, or urea derivatives in the presence of a reducing agent such as

dithiothreitol at alkaline pH or tris carboxyethyl phosphine at acid pH to release, break apart, and solubilize the inclusion bodies. The alpha 2 polypeptide in its now soluble form can then be analyzed using gel electrophoresis, immunoprecipitation, or the like. If it is desired to isolate the alpha 2 polypeptide, isolation (either as a monomer or in the form of a homo- or heterodimer) may be accomplished using standard methods such as those set forth below and in Marston et al., *Meth. Enz.*, 182:264-275 (1990).

In some cases, an alpha 2 polypeptide may not be biologically active upon isolation. Various methods for "refolding" or converting the polypeptide to its tertiary and/or quaternary structure and generating disulfide linkages, can be used to restore biological activity. Such methods include exposing the solubilized polypeptide to a pH usually above 7 and in the presence of a particular concentration of a chaotrope. The selection of chaotrope is very similar to the choices used for inclusion body solubilization, but usually the chaotrope is used at a lower concentration and is not necessarily the same as chaotropes used for the solubilization. In most cases, the refolding/oxidation solution will also contain a reducing agent or the reducing agent plus its oxidized form in a specific ratio to generate a particular redox potential allowing for disulfide shuffling to occur in the formation of the protein's cysteine bridge(s). Some of the commonly used redox couples include cysteine/cystamine, glutathione (GSH)/dithiobis GSH, cupric chloride, dithiothreitol(DTT)/dithiane DTT, and 2-mercaptoethanol(bME)/dithio-b(ME). In many instances, a cosolvent may be used or may be needed to increase the efficiency of the refolding and the more common

reagents used for this purpose include glycerol, polyethylene glycol of various molecular weights, arginine and the like.

If inclusion bodies are not formed to a significant degree upon expression of an alpha 2 polypeptide, the polypeptide will be found primarily in the supernatant after centrifugation of the cell homogenate and may be further isolated from the supernatant using methods such as those set forth below.

In situations where it is preferable to partially or completely purify an alpha 2 polypeptide such that it is partially or substantially free of contaminants, standard methods known to the one skilled in the art may be used. Such methods include, without limitation, separation by electrophoresis followed by electroelution, various types of chromatography (affinity, immunoaffinity, molecular sieve, and/or ion exchange), and/or high pressure liquid chromatography. In some cases, it may be preferable to use more than one of these methods for complete purification.

Alpha 2 polypeptides, fragments, and/or derivatives thereof may also be prepared by chemical synthesis methods (such as solid phase peptide synthesis) using techniques known in the art such as those set forth by Merrifield et al., *J. Am. Chem. Soc.*, 85:2149 (1963), Houghten et al. (*Proc Natl Acad. Sci. USA*, 82:5132 (1985), and Stewart and Young, Solid Phase Peptide Synthesis, Pierce Chemical Co., Rockford, IL (1984). Such polypeptides may be synthesized with or without a methionine on the amino terminus. Chemically synthesized alpha 2 polypeptides or fragments may be oxidized using methods set forth

in these references to form disulfide bridges. Alpha 2 polypeptides, fragments or derivatives are expected to have comparable biological activity to the corresponding alpha 2 polypeptides, fragments or derivatives produced recombinantly or purified from natural sources, and thus may be used interchangeably with recombinant or natural alpha 2 polypeptide.

Another means of obtaining alpha 2 polypeptide is via purification from biological samples such as source tissues and/or fluids in which the alpha 2 polypeptide is naturally found. Such purification can be conducted using methods for protein purification as described above. The presence of the alpha 2 polypeptide during purification may be monitored using, for example, an antibody prepared against recombinantly produced alpha 2 polypeptide or peptide fragments thereof.

Polypeptides

Polypeptides of the invention include isolated alpha 2 polypeptides (either monomers, homodimers or heterodimers) and polypeptides related thereto including fragments, variants, fusion polypeptides, and derivatives as defined herein.

Preferred alpha 2 polypeptide variants include glycosylation variants wherein the number and/or type of glycosylation sites has been altered compared to native alpha 2 polypeptide. In one embodiment, alpha 2 variants comprise a greater or a lesser number of N-linked glycosylation sites. An N-linked glycosylation site is characterized by the sequence: Asn-X-Ser or Thr, where the amino acid residue designated as X may be any type of amino acid except proline. Substitution(s) of amino acid

residues to create this sequence provides a potential new site for addition of an N-linked carbohydrate chain. Alternatively, substitutions to eliminate this sequence will remove an existing N-linked carbohydrate chain. Also provided is a rearrangement of N-linked carbohydrate chains wherein one or more N-linked glycosylation sites (typically those that are naturally occurring) are eliminated and one or more new N-linked sites are created. Additional preferred alpha 2 variants include cysteine variants, wherein one or more cysteine residues are deleted or substituted with another amino acid (e.g., serine). Cysteine variants are useful when alpha 2 must be refolded into a biologically active conformation such as after isolation of insoluble inclusive bodies. Cysteine variants generally have fewer cysteine residues than the native protein, and typically have an even number to minimize interactions resulting from unpaired cysteines.

One skilled in the art will be able to determine suitable variants of the native alpha 2 polypeptide using well known techniques. For example, one may be able to predict suitable areas of the molecule that may be changed without destroying biological activity. Also, one skilled in the art will realize that even areas that may be important for biological activity or for structure may be subject to conservative amino acid substitutions without destroying the biological activity or without adversely affecting the polypeptide structure.

For predicting suitable areas of the molecule that may be changed without destroying activity, one skilled in the art may target areas not believed to be important for activity. For example, when similar polypeptides with similar activities from

the same species or from other species are known, one skilled in the art may compare the amino acid sequence of alpha 2 polypeptide to such similar polypeptides. After making such a comparison, one skilled in the art would be able to determine residues and portions of the molecules that are conserved among similar polypeptides. One skilled in the art would know that changes in areas of the alpha 2 molecule that are not conserved would be less likely to adversely affect biological activity and/or structure. One skilled in the art would also know that, even in relatively conserved regions, one could have likely substituted chemically similar amino acids for the naturally occurring residues while retaining activity (conservative amino acid residue substitutions).

Also, one may review structure-function studies identifying residues in similar polypeptides that are important for activity or structure. In view of such a comparison, one skilled in the art can predict the importance of amino acid residues in alpha 2 that correspond to amino acid residues that are important for activity or structure in similar polypeptides. One may opt for chemically similar amino acid substitutions for such predicted important amino acid residues of alpha 2.

If available, one can also analyze the three-dimensional structure and amino acid sequence in relation to that structure in similar polypeptides. In view of that information, one skilled in the art may be able to predict the alignment of amino acid residues of alpha 2 polypeptide with respect to its three dimensional structure. One may choose not to make radical changes to amino acid residues predicted to be on the surface of the protein, since such

residues may be involved in important interactions with other molecules.

Moreover, one could generate test variants containing a single amino acid substitution at each amino acid residue. The variants could be screened for activity. Such variants could be used to gather information about suitable variants. For example, if one discovered that a change to a particular amino acid residue resulted in destroyed activity, variants with such a change would be avoided. In other words, based on information gathered from such experiments, when attempting to find additional acceptable variants, one skilled in the art would have known the amino acids where further substitutions should be avoided either alone or in combination with other mutations.

Alpha 2 fusion polypeptides of the invention comprise alpha 2 polypeptides, fragments, variants, or derivatives fused to a heterologous peptide or protein. Heterologous peptides and proteins include, but are not limited to: an epitope to allow for detection and/or isolation of an alpha 2 fusion polypeptide; a transmembrane receptor protein or a portion thereof, such as an extracellular domain, or a transmembrane and intracellular domain; a ligand or a portion thereof which binds to a transmembrane receptor protein; an enzyme or portion thereof which is catalytically active; a protein or peptide which promotes oligomerization, such as leucine zipper domain; and a protein or peptide which increases stability, such as an immunoglobulin constant region. An alpha 2 polypeptide may be fused to itself or to a fragment, variant, or derivative thereof. Fusions may be made either at the amino terminus or at the carboxy terminus of an alpha 2 polypeptide, and may be direct

with no linker or adapter molecule or may be through a linker or adapter molecule, such as one or more amino acid residues up to about twenty amino acids residues, or up to about fifty amino acid residues. A linker or adapter molecule may also be designed with a cleavage site for a DNA restriction endonuclease or for a protease to allow for separation of the fused moieties.

In a further embodiment of the invention, an alpha 2 polypeptide, fragment, variant and/or derivative is fused to an Fc region of human IgG. In one example, a human IgG hinge, CH2 and CH3 region may be fused at either the N-terminus or C-terminus of the alpha 2 polypeptides using methods known to the skilled artisan. In another example, a portion of a hinge regions and CH2 and CH3 regions may be fused. The alpha 2 Fc-fusion polypeptide so produced may be purified by use of a Protein A affinity column. In addition, peptides and proteins fused to an Fc region have been found to exhibit a substantially greater half-life *in vivo* than the unfused counterpart. Also, a fusion to an Fc region allows for dimerization/multimerization of the fusion polypeptide. The Fc region may be a naturally occurring Fc region, or may be altered to improve certain qualities, such as therapeutic qualities, circulation time, reduce aggregation, etc.

Alpha 2 polypeptide derivatives are also included in the scope of the present invention. Such derivatives are chemically modified alpha 2 polypeptide compositions in which alpha 2 polypeptide is linked to a polymer. The polymer selected is typically water soluble so that the protein to which it is attached does not precipitate in an aqueous environment, such as a physiological environment. The

polymer may be of any molecular weight, and may be branched or unbranched. Included within the scope of alpha 2 polypeptide polymers is a mixture of polymers. Preferably, for therapeutic use of the end-product preparation, the polymer will be pharmaceutically acceptable.

The water soluble polymer or mixture thereof may be selected from the group consisting of, for example, polyethylene glycol (PEG), monomethoxy-polyethylene glycol, dextran (such as low molecular weight dextran, of, for example about 6 kD), cellulose, or other carbohydrate based polymers, poly-(N-vinyl pyrrolidone) polyethylene glycol, propylene glycol homopolymers, a polypropylene oxide/ethylene oxide co-polymer, polyoxyethylated polyols (e.g., glycerol) and polyvinyl alcohol. Also encompassed by the invention are bifunctional PEG crosslinking molecules which may be used to prepare covalently attached alpha 2 multimers.

For the acylation reactions, the polymer(s) selected should have a single reactive ester group. For reductive alkylation, the polymer(s) selected should have a single reactive aldehyde group. A reactive aldehyde is, for example, polyethylene glycol propionaldehyde, which is water stable, or mono C1-C10 alkoxy or aryloxy derivatives thereof (see U.S. Patent No. 5,252,714).

Pegylation of alpha 2 polypeptides may be carried out by any of the pegylation reactions known in the art, as described for example in the following references: Francis et al., Focus on Growth Factors 3, 4-10 (1992); EP 0 154 316; EP 0 401 384 and U.S. Patent No. 4,179,337. Pegylation may be carried out via an acylation reaction or an alkylation reaction

with a reactive polyethylene glycol molecule (or an analogous reactive water-soluble polymer) as described below.

One water-soluble polymer for use herein is polyethylene glycol, abbreviated as "PEG". As used herein, polyethylene glycol is meant to encompass any of the forms of PEG that have been used to derivatize other proteins, such as mono-(C1-C10) alkoxy- or aryloxy-polyethylene glycol.

In general, chemical derivatization may be performed under any suitable conditions used to react a biologically active substance with an activated polymer molecule. Methods for preparing pegylated alpha 2 polypeptides will generally comprise the steps of (a) reacting the polypeptide with polyethylene glycol (such as a reactive ester or aldehyde derivative of PEG) under conditions whereby alpha 2 polypeptide becomes attached to one or more PEG groups, and (b) obtaining the reaction product(s). In general, the optimal reaction conditions for the acylation reactions will be determined based on known parameters and the desired result. For example, the larger the ratio of PEG: protein, the greater the percentage of poly-pegylated product.

In a preferred embodiment, the alpha 2 polypeptide derivative will have a single PEG moiety at the amino terminus. See U.S. Patent No. 5,234,784, herein incorporated by reference.

Generally, conditions which may be alleviated or modulated by administration of the present alpha 2 polypeptide derivative include those described herein for alpha 2 polypeptides. However, the alpha 2 polypeptide derivative disclosed herein

may have additional activities, enhanced or reduced biological activity, or other characteristics, such as increased or decreased half-life, as compared to the non-derivatized molecules.

Antibodies

Alpha 2 polypeptides, fragments, variants and derivatives may be used to prepare antibodies using methods known in the art. Thus, antibodies and antibody fragments that bind alpha 2 polypeptides are within the scope of the present invention. Antibodies may be polyclonal monospecific polyclonal, monoclonal, recombinant, chimeric, humanized, fully human, single chain and/or bispecific.

Polyclonal antibodies directed toward an alpha 2 polypeptide generally are raised in animals (e.g., rabbits or mice) by multiple subcutaneous or intraperitoneal injections of alpha 2 and an adjuvant. It may be useful to conjugate an alpha 2 polypeptide, or a variant, fragment or derivative thereof to a carrier protein that is immunogenic in the species to be immunized, such as keyhole limpet heocyanin, serum, albumin, bovine thyroglobulin, or soybean trypsin inhibitor. Also, aggregating agents such as alum are used to enhance the immune response. After immunization, the animals are bled and the serum is assayed for anti-alpha 2 antibody titer.

Monoclonal antibodies directed toward alpha 2 are produced using any method which provides for the production of antibody molecules by continuous cell lines in culture. Examples of suitable methods for preparing monoclonal antibodies include hybridoma methods of Kohler, et al., Nature 256, 495-497 (1975), and the human B-cell hybridoma method, Kozbor, J.

Immunol. 133, 3001 (1984); Brodeur, et al., Monoclonal Antibody Production Techniques and Applications, pp. 51-63 (Marcel Dekker, Inc., New York, 1987).

Also provided by this invention are hybridoma cell lines which produce monoclonal antibodies reactive with alpha 2 polypeptides.

Monoclonal antibodies of the invention may be modified for use as therapeutics. One embodiment is a "chimeric" antibody in which a portion of the heavy and/or light chain is identical with or homologous to corresponding sequence in antibodies derived from a particular species or belonging to a particular antibody class or subclass, while the remainder of the chain(s) is identical with or homologous to corresponding sequence in antibodies derived from another species or belonging to another antibody class or subclass, as well as fragments of such antibodies, so long as they exhibit the desired biological activity (see U.S. Patent No. 4,816,567; Morrison, et al., Proc. Natl. Acad. Sci. 81, 6851-6855 (1985)).

In another embodiment, a monoclonal antibody of the invention is a "humanized" antibody. Methods for humanizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. Humanization can be performed following methods known in the art (Jones, et al., Nature 321, 522-525 (1986); Riechmann, et al., Nature, 332, 323-327 (1988); Verhoeyen, et al., Science 239, 1534-1536 (1988)), by substituting rodent complementarily-determining regions (CDRs) for the corresponding regions of a human antibody.

Also encompassed by the invention are fully human antibodies which bind alpha 2 polypeptides, fragments, variants and/or derivatives. Such antibodies are produced by immunization with an alpha 2 antigen (optionally conjugated to a carrier) of transgenic animals (e.g., mice) that are capable of producing a repertoire of human antibodies in the absence of endogenous immunoglobulin production. See, for example, Jakobovits, et al., Proc. Natl. Acad. Sci. 90, 2551-2555 (1993); Jakobovits, et al., Nature 362, 255-258 (1993); Bruggermann, et al., Year in Immuno. 7,33 (1993). Human antibodies can also be produced in phage-display libraries (Hoogenboom, et al., J. Mol. Biol. 227, 381 (1991); Marks, et al., J. Mol. Biol. 222, 581 (1991).

Chimeric, CDR grafted and humanized antibodies are typically produced by recombinant methods. Nucleic acid encoding the antibodies are introduced into host cells and expressed using materials and procedures described herein above. In a preferred embodiment, the antibodies are produced in mammalian host cells, such as CHO cells. Fully human antibodies may be produced by expression of recombinant DNA transfected into host cells or by expression in hybridoma cells as described above.

For diagnostic applications, in certain embodiments, anti-alpha 2 antibodies typically will be labeled with a detectable moiety. The detectable moiety can be any one which is capable of producing, either directly or indirectly, a detectable signal. For example, the detectable moiety may be a radioisotope, such as ^{3}H, ^{14}C, ^{32}P, ^{35}S, or ^{125}I, a fluorescent or chemiluminescent compound, such as fluorescein isothiocyanate, rhodamine, or luciferin; or an enzyme, such as alkaline phosphatase, β-

galactosidase or horseradish peroxidase. Bayer, et al., Meth. Enz. 184: 138-163 (1990).

The anti-alpha 2 antibodies of the invention may be employed in any known assay method, such as competitive binding assays, direct and indirect sandwich assays, and immunoprecipitation assays (Sola, Monoclonal Antibodies: A Manual of Techniques, pp. 147-158 (CRC Press, Inc., 1987)) for detection and quantitation of alpha 2 polypeptides. The antibodies will bind alpha 2 polypeptides with an affinity which is appropriate for the assay method being employed.

Competitive binding assays rely on the ability of a labeled standard (e.g., an alpha 2 polypeptide, or an immunologically reactive portion thereof) to compete with the test sample analyte (an alpha 2 polypeptide) for binding with a limited amount of anti-alpha 2 antibody. The amount of an alpha 2 polypeptide in the test sample is inversely proportional to the amount of standard that becomes bound to the antibodies. To facilitate determining the amount of standard that becomes bound, the antibodies typically are insolubilized before or after the competition, so that the standard and analyte that are bound to the antibodies may conveniently be separated from the standard and analyte which remain unbound.

Sandwich assays involve the use of two antibodies, each capable of binding to a different immunogenic portion, or epitope, of the protein to be detected and/or quantitated. In a sandwich assay, the test sample analyte is typically bound by a first antibody which is immobilized on a solid support, and thereafter a second antibody binds to the analyte, thus forming an insoluble three part complex. See,

for example, U.S. Patent No. 4,376,110. The second antibody may itself be labeled with a detectable moiety (direct sandwich assays) or may be measured using an anti-immunoglobulin antibody that is labeled with a detectable moiety (indirect sandwich assays). For example, one type of sandwich assay is an ELISA assay, in which case the detectable moiety is an enzyme.

The anti-alpha 2 antibodies of the invention also are useful for *in vivo* imaging, wherein an antibody labeled with a detectable moiety is administered to an animal, preferably into the bloodstream, and the presence and location of the labeled antibody in the host is assayed. The antibody may be labeled with any moiety that is detectable in an animal, whether by nuclear magnetic resonance, radiology, or other detection means known in the art.

The invention also relates to a kit comprising anti-alpha 2 antibodies and other reagents useful for detecting alpha 2 levels in biological samples. Such reagents may include a secondary activity, a detectable label, blocking serum, positive and negative control samples and detection reagents.

Practical Applications

As mentioned above, the α2 polypeptide of this invention can be used to clone α2 receptors, using an "expression cloning" strategy. Radiolabeled (125-Iodine) α2 or "affinity/activity-tagged" α2 (such as an Fc fusion or an alkaline phosphatase fusion) can be used in binding assays to identify a cell type or cell line or tissue that expresses α2 receptor(s). RNA isolated from such cells or tissues would be converted to cDNA, cloned into a mammalian expression

vector, and transfected into mammalian cells (for example, COS, 293) to create an expression library. Radiolabeled or tagged α2 would then be used as an affinity ligand to identify and isolate the subset of cells in this library expressing the α2 receptor(s) on their surface. DNA would be isolated from these cells and transfected into mammalian cells to create a secondary expression library in which the fraction of cells expressing α2 receptor(s) would be many-fold higher than in the original library. This enrichment process would be repeated iteratively until a single recombinant clone containing an α2 receptor is isolated. Isolation of the α2 receptor(s) would be very useful in terms of being able to identify or develop novel agonists and antagonists of the α2 signaling pathway. Such agonists and antagonists would include soluble α2 receptor(s), anti-α2-receptor antibodies, small molecules or antisense oligonucleotides, and they could be used for treating one or more of the diseases/disorders listed below.

Also mentioned above is the use of α2 polypeptide for determining whether heterodimerization might occur between α2 and any of the five known glycoprotein hormone polypeptides (α, TSH-β, FSH-β, LH-β, and CG-β). This would be done by co-transfecting mammalian cells (COS, 293) with a vector capable of expressing His-tagged α2 (α2-His) and a vector capable of expressing a FLAG-tagged known glycoprotein hormone polypeptide (for example FSH-β-FLAG); Furuhashi et al., Molecular Endocrinology, Volume 9, pages 54-63 (1995). Conditioned media would be harvested 24-48 hours after transfection and immunoprecipitation would be carried out using a commercially available monoclonal antibody directed against the His-tag. The resulting immunoprecipitate would be subjected to denaturing polyacrylamide gel

electrophoresis, blotted onto a nylon membrane and probed (via Western blot) with a commercially available monoclonal antibody directed against the FLAG-tag. If the FLAG-tagged known glycoprotein hormone polypeptide is present on the Western blot this would indicate that it had heterodimerized (and thus immunoprecipitated) with α2. These co-transfections with α2-His would be done separately with each of the known glycoprotein hormone polypeptides. Formation of heterodimers for any given pair would indicate the potential for such a heterodimer to exist *in vivo* and would represent a possible α2 heterodimeric hormone. Recombinant α2 heterodimeric hormone (without tags) would be produced in mammalian cells, injected into a mammal (for example, mice) and assessed for *in vivo* function and utility in the diagnosis and/or treatment of one or more of the diseases or disorders listed below.

Biological function is anticipated for the α2 polypeptides of this invention similar to those of the glycoprotein hormones FAS, TSH, FSH, LH and CG, which, among other things, are known to act as growth factors in promoting the development (proliferation, differentiation) of prolactin producing cells, the thyroid gland and the gonads. These glycoproteins also act as endocrine hormones in their role as regulators of placental, thyroidal and gonadal function. FAS plays a role in stimulating prolactin secretion from decidual cells in the placenta, TSH plays a major role in the regulation of basal metabolism via the thyroid gland, and FSH, LH and CG play critical roles in male and female fertility, as well as in pregnancy. As such, hormones containing the α2 polypeptides of this invention (whether monomeric, homodimeric or heterodimeric) may also play

roles in the regulation of basal metabolism, fertility and pregnancy.

It has been determined by Northern analysis and *in situ* hybridization that the α2 polypeptide of this invention is expressed in the following organs or tissues: anterior pituitary, placenta, pancreas, adrenal cortex, intestinal crypts and gall bladder mucosa. In terms of potential function, the fact that α2 is expressed in many of the organs and tissues that make up the endocrine system suggests an important role for α2 hormones in the regulation and coordination of various endocrine system functions. The expression of α2 in pituitary, pancreas, adrenal cortex, intestinal crypts and gall bladder tissues indicates a possible role for α2 hormones in the shared function of these five organs or tissues, namely, energy/nutritional homeostasis (i.e., energy balance, basal metabolic rate, digestion, glucose homeostasis, distribution of body fat, general growth).

In addition, the expression of α2 in the pituitary and adrenal glands indicates a possible role for α2 hormones in one of the critical functions subserved by these two important organs, namely, the body's ability to cope with a variety of environmental and physiological stresses (for example, infection, fever, inflammation, fasting, high and low blood pressure, anxiety, shock). The expression of α2 in the pituitary gland and the placenta indicates a possible role for α2 hormones in the shared function of these two organs, namely fertility/pregnancy.

Finally, α2-containing hormones may act as growth factors involved in the regeneration (proliferation and differentiation) of tissues and

specialized cell types present in the pituitary gland, the placenta, the pancreas, the adrenal gland, the intestine and the gall bladder.

Based on these various potential functions, a hormone consisting of, or containing, a polypeptide of this invention may be useful for treating one or more of the following diseases or disorders: obesity, wasting syndromes (for example, cancer associated cachexia), gastrointestinal disorders (for example, ulcers), diabetes, hypertension, immune system dysfunction [for example, excessive inflammation, susceptibility to infection (such as AIDS), poor wound healing, asthma, arthritis and allergies], shock, tissue damage/degeneration (such as caused by cancer treatments), cancers, hyperplasias/hypertrophies, infertility, fertility (contraception) and impotence.

Antibodies of the invention may also be used as therapeutics. Therapeutic antibodies are generally agonists or antagonists, in that they either enhance or reduce, respectively, at least one of the biological activities of an alpha 2 polypeptide. In one embodiment, antagonist antibodies of the invention are antibodies or binding fragments thereof which are capable of specifically binding to an alpha 2 polypeptide, fragment, variant and/or derivative, and which are capable of inhibiting or eliminating the functional activity of an alpha 2 polypeptide *in vivo* or *in vitro*. In preferred embodiments, an antagonist antibody will inhibit the functional activity of an alpha 2 polypeptide at least about 50%, and preferably at least about 80%. In another embodiment antagonist antibodies are capable of interacting with an alpha 2 binding partner thereby inhibiting or eliminating alpha 2 activity *in vitro* or *in vivo*. Agonist and

antagonist anti-alpha 2 antibodies are identified by screening assays described herein.

Transgenic and Knockout Animals

Additionally included within the scope of the present invention are non-human animals such as mice, rats, or other rodents, rabbits, goats, or sheep, or other farm animals, in which the gene (or genes) encoding a native alpha 2 polypeptide has (have) been disrupted ("knocked out") such that the level of expression of this gene or genes is (are) significantly decreased or completely abolished. Such animals, which may be highly useful in elucidating the biology of alpha 2, may be prepared using *in vivo* techniques and methods such as those described in U.S. Patent No. 5,557,032.

The present invention further includes non-human animals such as mice, rats, or other rodents, rabbits, goats, or sheep, or other farm animals, in which the gene (or genes) encoding alpha 2 polypeptides in which either the native form of the gene(s) for that animal or a heterologous alpha 2 polypeptide gene(s) is (are) over expressed by the animal, thereby creating a "transgenic" animal. Such transgenic animals may be prepared using well known methods such as those described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122.

The present invention further includes non-human animals in which the promoter for one or more of the alpha 2 polypeptides of the present invention is either activated or inactivated (e.g., by using homologous recombination methods as described below) to alter the level of expression of one or more of the native alpha 2 polypeptides.

Such non-human animals may be used for drug candidate screening. The impact of a drug candidate on the animal may be measured. For example, drug candidates may decrease or increase expression of the alpha 2 polypeptide gene. In certain embodiments, the amount of alpha 2 polypeptide or a fragment(s) that is produced may be measured after exposure of the animal to the drug candidate. In certain embodiments, one may detect the actual impact of the drug candidate on the animal. For example, overexpression of a particular gene may result in, or be associated with, a disease or pathological condition. In such cases, one may test a drug candidate's ability to decrease expression of the gene or its ability to prevent or inhibit a pathological condition. In other examples, production of a particular metabolic product such as a fragment of a polypeptide, may result in, or be associated with, a disease or pathological condition. In such cases, one may test a drug candidate's ability to decrease production of such a metabolic product or its ability to prevent or inhibit a pathological condition.

Modulators of Alpha 2 Polypeptide Activity

In some situations, it may be desirable to identify molecules that are modulators, *i.e.*, agonists or antagonists, of the activity of alpha 2 polypeptide.

Natural or synthetic molecules that modulate alpha 2 can be identified using one or more screening assays, such as those described below. Such molecules may be administered either in an *ex vivo* manner, or in an *in vivo* manner by local or iv injection, or by oral delivery, implantation device, or the like.

The following definition is used herein for describing the assays:

"Test molecule(s)" refers to the molecule(s) that is/are under evaluation for the ability to modulate (i.e., increase or decrease) the activity of an alpha 2 polypeptide. Most commonly, a test molecule will interact directly with an alpha 2 polypeptides. However, it is also contemplated that a test molecule may also modulate alpha 2 activity indirectly, such as by affecting alpha 2 gene expression, or by binding to an alpha 2 binding partner (e.g., receptor or ligand). In one embodiment, test molecule will bind to an alpha 2 polypeptide with an affinity constant of at least about 10^{-6} M, preferably about 10^{-8} M, more preferably about 10^{-9} M, and even more preferably about 10^{-10} M.

Methods for identifying compounds which interact with alpha 2 polypeptides are encompassed by the invention. In certain embodiments, an alpha 2 polypeptide is incubated with a test molecule under conditions which permit interaction of the test molecule with an alpha 2 polypeptide, and the extent of the interaction can be measured. The test molecule(s) may be screened in a substantially purified form or in a crude mixture. Test molecule(s) may be nucleic acid molecules, proteins, peptides, carbohydrates, lipids or small molecular weight organic or inorganic compounds. Once a set of test molecules has been identified as interacting with an alpha 2 polypeptide, the molecules may be further evaluated for their ability to increase or decrease alpha 2 activity.

Measurement of the interaction of test molecules with alpha 2 polypeptides may be carried out in several formats, including cell-based binding assays, membrane binding assays, solution-phase assays and immunoassays. In general, test molecules are incubated with an alpha 2 polypeptide for a specified period of time and alpha 2 activity is determined by one or more assays described herein for measuring biological activity.

Interaction of test molecules with alpha 2 polypeptides may also be assayed directly using polyclonal or monoclonal antibodies in an immunoassay. Alternatively, modified forms of alpha 2 polypeptides containing epitope tags as described above may be used in solution and immunoassays.

In certain embodiments, an alpha 2 agonist or antagonist may be a protein, peptide, carbohydrate, lipid or small molecular weight molecule which interacts with alpha 2 to regulate its activity. Potential protein antagonists of alpha 2 include antibodies which interest with active regions of the polypeptide and inhibit or eliminate at least one activity of alpha 2. Molecules which regulate alpha 2 polypeptide expression may include nucleic acids which are complementary to nucleic acids encoding an alpha 2 polypeptide, or are complementary to nucleic acids sequences which direct or control expression of alpha 2 polypeptide, and which act as anti-sense regulators of expression.

In the event that alpha 2 polypeptides display biological activity through interaction with a binding partner (e.g., a receptor or a ligand), a variety of *in vitro* assays may be used to measure binding of an alpha 2 polypeptide to a corresponding

binding partner. These assays may be used to screen test molecules for their ability to increase or decrease the rate and/or the extent of binding of an alpha 2 polypeptide to its binding partner. In one assay, an alpha 2 polypeptide is immobilized by attachment to the bottom of the wells of a microtiter plate. Radiolabeled alpha 2 binding partner (for example, iodinated alpha 2 binding partner) and the test molecule(s) can then be added either one at a time (in either order) or simultaneously to the wells. After incubation, the wells can be washed and counted using a scintillation counter for radioactivity to determine the extent of binding to alpha 2 protein by its binding partner. Typically, the molecules will be tested over a range of concentrations, and a series of control wells lacking one or more elements of the test assays can be used for accuracy in evaluation of the results. An alternative to this method involves reversing the "positions" of the proteins, *i.e.*, immobilizing alpha 2 binding partner to the microtiter plate wells, incubating with the test molecule and radiolabeled alpha 2, and determining the extent of alpha 2 binding; see, for example, chapter 18 of *Current Protocols in Molecular Biology*, Ausubel et al., eds., John Wiley & Sons, New York, NY (1995).

As an alternative to radiolabelling, an alpha 2 polypeptide or its binding partner may be conjugated to biotin and the presence of biotinylated protein can then be detected using streptavidin linked to an enzyme, such as horse radish peroxidase [HRP] or alkaline phosphatase [AP], that can be detected colorometrically, or by fluorescent tagging of streptavidin. An antibody directed to an alpha 2 polypeptide or to an alpha 2 binding partner and that is conjugated to biotin may also be used and can be

detected after incubation with enzyme-linked streptavidin linked to AP or HRP.

An alpha 2 polypeptide and an alpha 2 binding partner may also be immobilized by attachment to agarose beads, acrylic beads or other types of such inert substrates. The substrate-protein complex can be placed in a solution containing the complementary protein and the test compound; after incubation, the beads can be precipitated by centrifugation, and the amount of binding between an alpha 2 polypeptide and its binding partner can be assessed using the methods described above. Alternatively, the substrate-protein complex can be immobilized in a column and the test molecule and complementary protein passed over the column. Formation of a complex between an alpha 2 polypeptide and its binding partner can then be assessed using any of the techniques set forth above, *i.e.*, radiolabelling, antibody binding, or the like.

Another *in vitro* assay that is useful for identifying a test molecule which increases or decreases formation of a complex between an alpha 2 binding protein and an alpha 2 binding partner is a surface plasmon resonance detector system such as the Biacore assay system (Pharmacia, Piscataway, NJ). The Biacore system may be carried out using the manufacturer's protocol. This assay essentially involves covalent binding of either alpha 2 or an alpha 2 binding partner to a dextran-coated sensor chip which is located in a detector. The test compound and the other complementary protein can then be injected into the chamber containing the sensor chip either simultaneously or sequentially and the amount of complementary protein that binds can be assessed based on the change in molecular mass which is physically associated with the dextran-coated side

of the sensor chip; the change in molecular mass can be measured by the detector system.

In some cases, it may be desirable to evaluate two or more test compounds together for their ability to increase or decrease formation of a complex between an alpha 2 polypeptide and an alpha 2 binding partner complex. In these cases, the assays set forth above can be readily modified by adding such additional test compound(s) either simultaneous with, or subsequent to, the first test compound. The remainder of steps in the assay are as set forth above.

In vitro assays such as those described above may be used advantageously to screen rapidly large numbers of compounds for effects on complex formation by alpha 2 and alpha 2 binding partner. The assays may be automated to screen compounds generated in phage display, synthetic peptide and chemical synthesis libraries.

Compounds which increase or decrease formation of a complex between an alpha 2 polypeptide and an alpha 2 binding partner may also be screened in cell culture using cells and cell lines expressing either alpha 2 or alpha 2 binding partner. Cells and cell lines may be obtained from any mammal, but preferably will be from human or other primate, canine, or rodent sources. The binding of an alpha 2 polypeptide to cells expressing alpha 2 binding partner at the surface is evaluated in the presence or absence of test molecules and the extent of binding may be determined by, for example, flow cytometry using a biotinylated antibody to an alpha 2 binding partner. Cell culture assays may be used advantageously to further evaluate compounds that

score positive in protein binding assays described above.

Cell cultures can also be used to screen the impact of a drug candidate. For example, drug candidates may decrease or increase expression of the alpha 2 polypeptide gene. In certain embodiments, the amount of alpha 2 polypeptide or a fragment(s) that is produced may be measured after exposure of the cell culture to the drug candidate. In certain embodiments, one may detect the actual impact of the drug candidate on the cell culture. For example, overexpression of a particular gene may have a particular impact on the cell culture. In such cases, one may test a drug candidate's ability to increase or decrease expression of the gene or its ability to prevent or inhibit a particular impact on the cell culture. In other examples, production of a particular metabolic product such as a fragment of a polypeptide, may result in, or be associated with, a disease or pathological condition. In such cases, one may test a drug candidate's ability to decrease production of such a metabolic product in a cell culture.

A yeast two hybrid system can be used to identify novel polypeptides that bind to, or interact with, alpha 2 polypeptides; see Chien et al., Proc. Natl. Acad. Sci. USA 88, 9578-9583 (1991). As an example, hybrid constructs comprising DNA encoding a cytoplasmic domain of an alpha 2 polypeptide fused to a yeast GAL4-DNA binding domain may be used as a two-hybrid bait plasmid. Positive clones emerging from the screening may be characterized further to identify interacting proteins.

Pharmaceutical Compositions and Administration

Therapeutic compositions of alpha 2 polypeptides, which are also within the scope of the present invention, may comprise a therapeutically effective amount of an alpha 2 polypeptide, fragment, variant, or derivative in admixture with a pharmaceutically acceptable carrier and/or other pharmaceutically acceptable formulation agent(s). The carrier material may be water for injection, preferably supplemented with other materials common in solutions for administration to mammals. Typically, an alpha 2 polypeptide therapeutic compound will be administered in the form of a composition comprising purified polypeptide, fragment, variant, or derivative in conjunction with one or more physiologically acceptable agents. Neutral buffered saline or saline mixed with serum albumin are exemplary appropriate carriers. Preferably, the product is formulated as a lyophilizate using appropriate excipients (e.g., sucrose). Other standard pharmaceutically acceptable agents such as diluents and excipients may be included as desired. Other exemplary compositions comprise Tris buffer of about pH 7.0-8.5, or acetate buffer of about pH 4.0-5.5, which may further include sorbitol or a suitable substitute therefor.

Alpha 2 pharmaceutical compositions typically will include a therapeutically effective amount of alpha 2 protein in admixture with one or more pharmaceutically and physiologically acceptable formulation agents selected for suitability with the mode of administration. Suitable formulation agents include, but are not limited to, stabilizers, antioxidants, preservatives, coloring, flavoring and diluting agents, emulsifying agents, suspending agents, solubilizing agents (solvents), fillers,

bulking agents, buffers, delivery vehicles, diluents, excipients and/or pharmaceutical adjuvants. For example, a suitable vehicle may be water for injection, physiological saline solution, or artificial cerebrospinal fluid, possibly supplemented with other materials common in compositions for parenteral administration. Neutral buffered saline or saline mixed with serum albumin are further exemplary vehicles. The term "pharmaceutically acceptable carrier" or "physiologically acceptable carrier" as used herein refers to a formulation agent(s) suitable for accomplishing or enhancing the delivery of the alpha 2 protein as a pharmaceutical composition.

The primary solvent in a composition may be either aqueous or non-aqueous in nature. In addition, the vehicle may contain other formulation materials for modifying or maintaining the pH, osmolarity, viscosity, clarity, color, sterility, stability, rate of dissolution, or odor of the formulation. Similarly, the composition may contain additional formulation materials for modifying or maintaining the rate of release of alpha 2 protein, or for promoting the absorption or penetration of alpha 2 protein.

The alpha 2 polypeptide compositions can be administered parentally. Alternatively, the compositions may be administered intravenously or subcutaneously. When systemically administered, the therapeutic compositions for use in this invention may be in the form of a pyrogen-free, parentally acceptable aqueous solution. The preparation of such pharmaceutically acceptable protein solutions, with due regard to pH, isotonicity, stability and the like, is within the skill of the art.

Therapeutic formulations of alpha 2 polypeptide compositions useful for practicing the present invention may be prepared for storage by mixing the selected composition having the desired degree of purity with optional physiologically acceptable carriers, excipients, or stabilizers (*Remington's Pharmaceutical Sciences*, 18th Edition, A.R. Gennaro, ed., Mack Publishing Company [1990]) in the form of a lyophilized cake or an aqueous solution. Acceptable carriers, excipients or stabilizers preferably are nontoxic to recipients and are preferably inert at the dosages and concentrations employed, and preferably include buffers such as phosphate, citrate, or other organic acids; antioxidants such as ascorbic acid; low molecular weight polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, arginine or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; salt-forming counterions such as sodium; and/or nonionic surfactants such as Tween, pluronics or polyethylene glycol (PEG).

The optimal pharmaceutical formulation will be determined by one skilled in the art depending upon the intended route of administration, delivery format and desired dosage. See for example, Remington's Pharmaceutical Sciences, 18th Ed. (1990, Mack Publishing Co., Easton, PA 18042) pages 1435-1712. Such compositions may influence the physical state, stability, rate of *in vivo* release, and rate of *in vivo* clearance of the present alpha 2 protein.

An effective amount of an alpha 2 polypeptide composition to be employed therapeutically will depend, for example, upon the therapeutic objectives such as the indication for which the alpha 2 polypeptide is being used, the route of administration, and the condition of the patient. Accordingly, it may be necessary for the therapist to titer the dosage and modify the route of administration as required to obtain the optimal therapeutic effect. A typical dosage may range from about 0.1 µg/kg to up to about 100 mg/kg or more, depending on the factors mentioned above. In other embodiments, the dosage may range from 1 µg/kg up to about 100 mg/kg; or 5 µg/kg up to about 100 mg/kg; or 0.1 µg/kg up to about 100 mg/kg; or 1 µg/kg up to about 100 mg/kg Typically, a clinician will administer the composition until a dosage is reached that achieves the desired effect. The composition may therefore be administered as a single dose, or as two or more doses (which may or may not contain the same amount of alpha 2 polypeptide) over time, or as a continuous infusion via implantation device or catheter.

As further studies are conducted, information will emerge regarding appropriate dosage levels for treatment of various conditions in various patients, and the ordinary skilled worker, considering the therapeutic context, the type of disorder under treatment, the age and general health of the recipient, will be able to ascertain proper dosing.

The alpha 2 polypeptide composition to be used for *in vivo* parenteral administration typically must be sterile. This is readily accomplished by filtration through sterile filtration membranes. Where the composition is lyophilized, sterilization using these methods may be conducted either prior to,

or following, lyophilization and reconstitution. The composition for parenteral administration ordinarily will be stored in lyophilized form or in solution.

Therapeutic compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Effective administration forms, such as (1) slow-release formulations, (2) inhalant mists, or (3) orally active formulations are also envisioned. The alpha 2 pharmaceutical composition also may be formulated for parenteral administration. Such parenterally administered therapeutic compositions are typically in the form of a pyrogen-free, parenterally acceptable aqueous solution comprising alpha 2 in a pharmaceutically acceptable vehicle. The alpha 2 pharmaceutical compositions also may include particulate preparations of polymeric compounds such as polylactic acid, polyglycolic acid, etc., or the introduction of alpha 2 into liposomes. Hyaluronic acid may also be used, and this may have the effect of promoting sustained duration in the circulation.

A particularly suitable vehicle for parenteral injection is sterile distilled water in which alpha 2 is formulated as a sterile, isotonic solution, properly preserved. Yet another preparation may involve the formulation of alpha 2 with an agent, such as injectable microspheres, bio-erodible particles or beads, or liposomes, that provides for the controlled or sustained release of the protein product which may then be delivered as a depot injection. Other suitable means for the introduction

of alpha 2 include implantable drug delivery devices which contain the alpha 2.

The preparations of the present invention may include other components, for example parenterally acceptable preservatives, tonicity agents, cosolvents, wetting agents, complexing agents, buffering agents, antimicrobials, antioxidants and surfactants, as are well known in the art. For example, suitable tonicity enhancing agents include alkali metal halides (preferably sodium or potassium chloride), mannitol, sorbitol and the like. Suitable preservatives include, but are not limited to, benzalkonium chloride, thimerosal, phenethyl alcohol, methylparaben, propylparaben, chlorhexidine, sorbic acid and the like. Hydrogen peroxide may also be used as preservative. Suitable cosolvents are for example glycerin, propylene glycol and polyethylene glycol. Suitable complexing agents are for example caffeine, polyvinylpyrrolidone, beta-cyclodextrin or hydroxypropyl-beta-cyclodextrin. Suitable surfactants or wetting agents include sorbitan esters, polysorbates such as polysorbate 80, tromethamine, lecithin, cholesterol, tyloxapal and the like. The buffers can be conventional buffers such as borate, citrate, phosphate, bicarbonate, or Tris-HCl.

The formulation components are present in concentration that are acceptable to the site of administration. For example, buffers are used to maintain the composition at physiological pH or at slightly lower pH, typically within a pH range of from about 5 to about 8.

A pharmaceutical composition may be formulated for inhalation. For example, alpha 2 may be formulated as a dry powder for inhalation. alpha 2

inhalation solutions may also be formulated in a liquefied propellant for aerosol delivery. In yet another formulation, solutions may be nebulized.

It is also contemplated that certain formulations containing alpha 2 may be administered orally. Alpha 2 which is administered in this fashion may be formulated with or without those carriers customarily used in the compounding of solid dosage forms such as tablets and capsules. For example, a capsule may be designed to release the active portion of the formulation at the point in the gastrointestinal tract when bioavailability is maximized and pre-systemic degradation is minimized. Additional agents may be included to facilitate absorption of alpha 2. Diluents, flavorings, low melting point waxes, vegetable oils, lubricants, suspending agents, tablet disintegrating agents, and binders may also be employed.

Another preparation may involve an effective quantity of alpha 2 in a mixture with non-toxic excipients which are suitable for the manufacture of tablets. By dissolving the tablets in sterile water, or other appropriate vehicle, solutions can be prepared in unit dose form. Suitable excipients include, but are not limited to, inert diluents, such as calcium carbonate, sodium carbonate or bicarbonate, lactose, or calcium phosphate; or binding agents, such as starch, gelatin, or acacia; or lubricating agents such as magnesium stearate, stearic acid, or talc.

Additional alpha 2 formulations will be evident to those skilled in the art, including formulations involving alpha 2 in combination with one or more other therapeutic agents. Techniques for formulating a variety of other sustained- or

controlled-delivery means, such as liposome carriers, bio-erodible microparticles or porous beads and depot injections, are also known to those skilled in the art. See, for example, the Supersaxo et al. description of controlled release porous polymeric microparticles for the delivery of pharmaceutical compositions (See WO 93/15722 (PCT/US93/00829) the disclosure of which is hereby incorporated by reference.

Regardless of the manner of administration, the specific dose may be calculated according to body weight, body surface area or organ size. Further refinement of the calculations necessary to determine the appropriate dosage for treatment involving each of the above mentioned formulations is routinely made by those of ordinary skill in the art and is within the ambit of tasks routinely performed by them. Appropriate dosages may be ascertained through use of appropriate dose-response data.

The route of administration of the composition is in accordance with known methods, e.g. oral, injection or infusion by intravenous, intraperitoneal, intracerebral (intraparenchymal), intracerebroventricular, intramuscular, intraocular, intraarterial, or intralesional routes, or by sustained release systems or implantation device which may optionally involve the use of a catheter. Where desired, the compositions may be administered continuously by infusion, bolus injection or by implantation device.

One may further administer the present pharmaceutical compositions by pulmonary administration; see, for example, PCT WO94/20069, which discloses pulmonary delivery of chemically

modified proteins and is herein incorporated by reference. For pulmonary delivery, the particle size should be suitable for delivery to the distal lung. For example, the particle size may be from 1μm to 5μm, however, larger particles may be used if each particle is fairly porous.

Alternatively or additionally, the composition may be administered locally via implantation into the affected area of a membrane, sponge, or other appropriate material on to which alpha 2 polypeptide has been absorbed or encapsulated.

Where an implantation device is used, the device may be implanted into any suitable tissue or organ, and delivery of alpha 2 polypeptide may be directly through the device via bolus, or via continuous administration, or via catheter using continuous infusion.

Alpha 2 polypeptide may be administered in a sustained release formulation or preparation. Suitable examples of sustained-release preparations include semipermeable polymer matrices in the form of shaped articles, e.g., films or microcapsules. Sustained release matrices include polyesters, hydrogels, polylactides (U.S. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma ethyl-L-glutamate (Sidman *et al*, *Biopolymers*, 22: 547-556 [1983]), poly (2-hydroxyethyl-methacrylate) (Langer *et al.*, *J. Biomed. Mater. Res.*, 15: 167-277 [1981] and Langer, *Chem. Tech.*, 12: 98-105 [1982]), ethylene vinyl acetate (Langer *et al.*, *supra*) or poly-D(-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also may include liposomes, which can be prepared by any of several methods known in the art (*e.g.*, Eppstein *et al.*, *Proc. Natl. Acad. Sci. USA*,

82: 3688-3692 [1985]; EP 36,676; EP 88,046; EP 143,949).

The alpha 2 polypeptides, fragments thereof, variants and derivatives, may be employed alone, together, or in combination with other pharmaceutical compositions. The alpha 2 polypeptides, fragments, variants, and derivatives may be used in combination with cytokines, growth factors, antibiotics, anti-inflammatories, and/or chemotherapeutic agents as is appropriate for the indication being treated.

In some cases, it may be desirable to use alpha 2 polypeptide compositions in an *ex vivo* manner. Here, cells, tissues, or organs that have been removed from the patient are exposed to alpha 2 polypeptide compositions after which the cells, tissues and/or organs are subsequently implanted back into the patient.

Additional Practical Applications

An alpha 2 polypeptide of this invention may be delivered through implanting into patients certain cells that have been genetically engineered, using methods such as those described herein, to express and secrete the polypeptides, fragments, variants, or derivatives. Such cells may be animal or human cells, and may be derived from the patient's own tissue or from another source, either human or non-human. Optionally, the cells may be immortalized. However, in order to decrease the chance of an immunological response, it is preferred that the cells be encapsulated to avoid infiltration of surrounding tissues. The encapsulation materials are typically biocompatible, semi-permeable polymeric enclosures or membranes that allow release of the protein product(s)

but prevent destruction of the cells by the patient's immune system or by other detrimental factors from the surrounding tissues.

Methods used for membrane encapsulation of cells are familiar to the skilled artisan, and preparation of encapsulated cells and their implantation in patients may be accomplished without undue experimentation. See, for example, U.S. Patent Nos. 4,892,538; 5,011,472; and 5,106,627. A system for encapsulating living cells is described in PCT WO 91/10425 (Aebischer et al.). Techniques for formulating a variety of other sustained or controlled delivery means, such as liposome carriers, bio-erodible particles or beads, are also known to those in the art, and are described. The cells, with or without encapsulation, may be implanted into suitable body tissues or organs of the patient.

As discussed above, it may be desirable to treat isolated cell populations with one or more alpha 2 polypeptides, variants, derivatives and/or fragments directly.

Additional objects of the present invention relate to methods for both the *in vitro* production of therapeutic proteins by means of homologous recombination and for the production and delivery of therapeutic proteins by gene therapy.

It is further envisioned that alpha 2 protein may be produced by homologous recombination, or with recombinant production methods utilizing control elements introduced into cells already containing DNA encoding alpha 2. For example, homologous recombination methods may be used to modify a cell that contains a normally transcriptionally

silent alpha 2 gene, or under expressed gene, and thereby produce a cell which results in *de novo* or increased alpha 2 protein production from the cell's endogenous alpha 2 gene. Homologous recombination is a technique originally developed for targeting genes to induce or correct mutations in transcriptionally active genes (Kucherlapati, Prog. in Nucl. Acid Res. and Mol. Biol., 36:301, 1989). The basic technique was developed as a method for introducing specific mutations into specific regions of the mammalian genome (Thomas et al., Cell, 44:419-428, 1986; Thomas and Capecchi, Cell, 51:503-512, 1987; Doetschman et al., Proc. Natl. Acad. Sci., 85:8583-8587, 1988) or to correct specific mutations within defective genes (Doetschman et al., Nature, 330:576-578, 1987). Exemplary homologous recombination techniques are described in U.S. 5,272,071 (EP 91 90 3051, EP Publication No. 505 500; PCT/US90/07642, International Publication No. WO 91/09955).

Through homologous recombination, the DNA sequence to be inserted into the genome can be directed to a specific region of the gene of interest by attaching it to targeting DNA. The targeting DNA is a nucleotide sequence that is complementary (homologous) to a region of the genomic DNA. Small pieces of targeting DNA that are complementary to a specific region of the genome are put in contact with the parental strand during the DNA replication process. It is a general property of DNA that has been inserted into a cell to hybridize, and therefore, recombine with other pieces of endogenous DNA through shared homologous regions. If this complementary strand is attached to an oligonucleotide that contains a mutation or a different sequence or an additional nucleotide, it too is incorporated into the newly synthesized strand as a result of the recombination.

As a result of the proofreading function, it is possible for the new sequence of DNA to serve as the template. Thus, the transferred DNA is incorporated into the genome.

Attached to these pieces of targeting DNA are regions of DNA which may increase the expression of alpha 2 protein. For example, a promoter/enhancer element, a suppressor, or an exogenous transcription modulatory element is inserted in the genome of the intended host cell in proximity and orientation sufficient to influence the transcription of DNA encoding the desired alpha 2 protein. The control element controls a portion of the DNA present in the host cell genome. Thus, the expression of alpha 2 protein may be achieved not by transfection of DNA that encodes the alpha 2 gene itself, but rather by the use of targeting DNA (containing regions of homology with the endogenous gene of interest) coupled with DNA regulatory segments that provide the endogenous gene sequence with recognizable signals for production of alpha 2 protein.

In an exemplary method, expression of a desired targeted gene in a cell (i.e., a desired endogenous cellular gene) is altered by the introduction, by homologous recombination into the cellular genome at a preselected site, of DNA which includes at least a regulatory sequence, an exon and a splice donor site. These components are introduced into the chromosomal (genomic) DNA in such a manner that this, in effect, results in production of a new transcription unit (in which the regulatory sequence, the exon and the splice donor site present in the DNA construct are operatively linked to the endogenous gene). As a result of introduction of these

components into the chromosomal DNA, the expression of the desired endogenous gene is altered.

The present invention further relates to DNA constructs useful in the method of altering expression of a target gene. In certain embodiments, the exemplary DNA constructs comprise: (a) a targeting sequence; (b) a regulatory sequence; (c) an exon; and (d) an unpaired splice-donor site. The targeting sequence in the DNA construct directs the integration of elements (a)-(d) into a target gene in a cell such that the elements (b)-(d) are operatively linked to sequences of the endogenous target gene. In another embodiment, the DNA constructs comprise: (a) a targeting sequence, (b) a regulatory sequence, (c) an exon, (d) a splice-donor site, (e) an intron, and (f) a splice-acceptor site, wherein the targeting sequence directs the integration of elements (a)-(f) such that the elements of (b)-(f) are operatively linked to the endogenous gene. The targeting sequence is homologous to the preselected site in the cellular chromosomal DNA with which homologous recombination is to occur. In the construct, the exon is generally 3' of the regulatory sequence and the splice-donor site is 3' of the exon.

If the sequence of a particular gene is known, a piece of DNA that is complementary to a selected region of the gene can be synthesized or otherwise obtained, such as by appropriate restriction of the native DNA at specific recognition sites bounding the region of interest. This piece serves as a targeting sequence upon insertion into the cell and will hybridize to its homologous region within the genome. If this hybridization occurs during DNA replication, this piece of DNA, and any additional sequence attached thereto, will act as an Okazaki

fragment and will be backstitched into the newly synthesized daughter strand of DNA. The present invention, therefore, includes nucleotides encoding an alpha 2 molecule, which nucleotides may be used as targeting sequences.

An additional method by which homologous recombination may be used to increase, or cause, alpha 2 protein production from a cell's endogenous alpha 2 gene involves first using homologous recombination to place a recombination sequence from a site specific recombination system (e.g., Cre/loxP, FLP/FRT) (Sauer, Current Opinion In Biotechnology, 5: 521-527, 1994; Sauer, Methods In Enzymology, 225: 890-900, 1993) upstream (5'to) of the cell's endogenous genomic alpha 2 coding region. A plasmid containing a recombination site homologous to the site that was placed just upstream of the genomic alpha 2 coding region is introduced into the modified cell line along with the appropriate recombinase enzyme. This recombinase causes the plasmid to integrate, via the plasmid's recombination site, into the recombination site located just upstream of the genomic alpha 2 coding region in said cell line (Baubonis and Sauer, Nucleic Acids Research, 21: 2025-2029, 1993; O'Gorman et al., Science, 251: 1351-1355, 1991). Any sequences known to increase transcription (e.g., enhancer/promoter, intron, translational enhancer), if properly positioned in this plasmid, would integrate in such a manner as to create a new or modified transcriptional unit resulting in *de novo* or increased alpha 2 protein production from the cell's endogenous alpha 2 gene.

A further method to use the cell line in which the site specific recombination sequence had been placed just upstream of the cell's endogenous genomic alpha 2 coding region would be to use

homologous (or illegitimate) recombination to introduce a second recombination site elsewhere in the said cell line's genome. The appropriate recombinase enzyme would then be introduced into the two-recombination-site cell line and would cause a recombination event (deletion, inversion, translocation) (Sauer, Current Opinion In Biotechnology, 5: 521-527, 1994; Sauer, Methods In Enzymology, 225: 890-900, 1993) that would create a new or modified transcriptional unit resulting in *de novo* or increased alpha 2 protein production from the cell's endogenous alpha 2 gene.

An additional approach for increasing, or causing, expression of alpha 2 from a cell's endogenous alpha 2 gene involves increasing, or causing, the expression of a gene or genes (e.g., transcription factors) and/or decreasing the expression of a gene or genes (e.g., transcriptional repressors) in a manner which results in *de novo* or increased alpha 2 protein production from the cell's endogenous alpha 2 gene. This method includes the introduction of a non-naturally occurring polypeptide (e.g., a polypeptide comprising a site specific DNA binding domain fused to a transcriptional factor domain) into the cell such that *de novo* or increased alpha 2 protein production from the cell's endogenous alpha 2 gene results.

Alpha 2 cell therapy, such as the implantation of cells producing alpha 2, is also contemplated. This embodiment would involve implanting into patients cells capable of synthesizing and secreting a biologically active form of alpha 2. Such alpha 2-producing cells may be cells that are natural producers of alpha 2 or may be recombinant cells whose ability to produce alpha 2 has been

augmented by transformation with a gene encoding the desired alpha 2 molecule or with a gene augmenting the expression of alpha 2. Such a modification may be accomplished by means of a vector suitable for delivering the gene as well as promoting its expression and secretion. In order to minimize a potential immunological reaction in patients being administered an alpha 2 protein or polypeptide of a foreign species, it is preferred that the natural cells producing alpha 2 be of human origin and produce human alpha 2. Likewise, it is preferred that the recombinant cells producing alpha 2 be transformed with an expression vector containing a gene encoding a human alpha 2 molecule.

Implanted cells may be encapsulated to avoid infiltration of surrounding tissue. Human or non-human animal cells may be implanted in patients in biocompatible, semipermeable polymeric enclosures or membranes that allow release of alpha 2, but that prevent destruction of the cells by the patient's immune system or by other detrimental factors from the surrounding tissue. Alternatively, the patient's own cells, transformed to produce alpha 2 ex vivo, could be implanted directly into the patient without such encapsulation.

Techniques for the encapsulation of living cells are known in the art, and the preparation of the encapsulated cells and their implantation in patients may be accomplished without undue experimentation. For example, Baetge et al. (WO 95/05452; PCT/US94/09299, the disclosure of which is hereby incorporated by reference) describe membrane capsules containing genetically engineered cells for the effective delivery of biologically active molecules. The capsules are biocompatible and are easily

retrievable. The capsules encapsulate cells transfected with recombinant DNA molecules comprising DNA sequences coding for biologically active molecules operatively linked to promoters that are not subject to down regulation *in vivo* upon implantation into a mammalian host. The devices provide for the delivery of the molecules from living cells to specific sites within a recipient. In addition, see U.S. Patent Numbers 4,892,538, 5,011,472, and 5,106,627. A system for encapsulating living cells is described in PCT Application WO 91/10425 of Aebischer et al. See also, PCT Application WO 91/10470 of Aebischer et al., Winn et al., Exper. Neurol., 113:322-329, 1991, Aebischer et al., Exper. Neurol., 111:269-275, 1991; and Tresco et al., ASAIO, 38:17-23, 1992.

In vivo and *in vitro* gene therapy delivery of alpha 2 is also envisioned. *In vivo* gene therapy may be accomplished by introducing the gene encoding alpha 2 into cells via local injection of a polynucleotide molecule or other appropriate delivery vectors. (Hefti, J. Neurobiology,. 25:1418-1435, 1994). For example, a polynucleotide molecule encoding alpha 2 may be contained in an adeno-associated virus vector for delivery to the targeted cells (e.g., Johnson, International Publication No. WO 95/34670; International Application No. PCT/US95/07178). The recombinant adeno-associated virus (AAV) genome typically contains AAV inverted terminal repeats flanking a DNA sequence encoding alpha 2 operably linked to functional promoter and polyadenylation sequences.

Alternative viral vectors include, but are not limited to, retrovirus, adenovirus, herpes simplex virus and papilloma virus vectors. U.S. Patent No. 5,672,344 describes an *in vivo* viral-mediated gene

transfer system involving a recombinant neurotrophic HSV-1 vector. U.S. Patent No. 5,399,346 provides examples of a process for providing a patient with a therapeutic protein by the delivery of human cells which have been treated *in vitro* to insert a DNA segment encoding a therapeutic protein. Additional methods and materials for the practice of gene therapy techniques are described in U.S. Patent No. 5,631,236 involving adenoviral vectors; U.S. Patent No. 5,672,510 involving retroviral vectors; and U.S. 5,635,399 involving retroviral vectors expressing cytokines.

Nonviral delivery methods include liposome-mediated transfer, naked DNA delivery (direct injection), receptor-mediated transfer (ligand-DNA complex), electroporation, calcium phosphate precipitation and microparticle bombardment (e.g., gene gun). Gene therapy materials and methods may also include inducible promoters, tissue-specific enhancer-promoters, DNA sequences designed for site-specific integration, DNA sequences capable of providing a selective advantage over the parent cell, labels to identify transformed cells, negative selection systems and expression control systems (safety measures), cell-specific binding agents (for cell targeting), cell-specific internalization factors, transcription factors to enhance expression by a vector as well as methods of vector manufacture. Such additional methods and materials for the practice of gene therapy techniques are described in U.S. Patent No. 4,970,154 electroporation techniques; WO 96/40958 nuclear ligands; U.S. Patent No. 5,679,559 concerning a lipoprotein-containing system for gene delivery; U.S. Patent No. 5,676,954 involving liposome carriers; U.S. Patent No. 5,593,875 concerning methods for calcium phosphate transfection; and U.S. Patent

No. 4,945,050 wherein biologically active particles are propelled at cells at a speed whereby the particles penetrate the surface of the cells and become incorporated into the interior of the cells. Expression control techniques include chemical induced regulation (e.g., WO 96/41865 and WO 97/31899), the use of a progesterone antagonist in a modified steroid hormone receptor system (e.g., U.S. Patent No. 5,364,791), ecdysone control systems (e.g., WO 96/37609), and positive tetracycline-controllable transactivators (e.g., U.S. Patent No. 5,589,362; U.S. Patent No. 5,650,298; and U.S. Patent No. 5,654,168).

It is also contemplated that alpha 2 gene therapy or cell therapy can further include the delivery of a second protein. For example, the host cell may be modified to express and release both alpha 2 and a second protein. Alternatively, the alpha 2 and a second protein may be expressed in and released from separate cells. Such cells may be separately introduced into the patient or the cells may be contained in a single implantable device, such as the encapsulating membrane described above.

One manner in which gene therapy can be applied is to use the alpha 2 gene (either genomic DNA, cDNA, and/or synthetic DNA encoding an alpha 2 polypeptide, or a fragment, variant, or derivative thereof) which may be operably linked to a constitutive or inducible promoter to form a "gene therapy DNA construct". The promoter may be homologous or heterologous to the endogenous alpha 2 gene, provided that it is active in the cell or tissue type into which the construct will be inserted. Other components of the gene therapy DNA construct may optionally include, DNA molecules designed for site-specific integration (*e.g.*, endogenous flanking

sequences useful for homologous recombination), tissue-specific promoter, enhancer(s) or silencer(s), DNA molecules capable of providing a selective advantage over the parent cell, DNA molecules useful as labels to identify transformed cells, negative selection systems, cell specific binding agents (as, for example, for cell targeting), cell-specific internalization factors, and transcription factors to enhance expression by a vector as well as factors to enable vector manufacture.

This gene therapy DNA construct can then be introduced into the patient's cells (either *ex vivo* or *in vivo*). One means for introducing the gene therapy DNA construct is via viral vectors. Suitable viral vectors typically used in gene therapy for delivery of gene therapy DNA constructs include, without limitation, adenovirus, adeno-associated virus, herpes simplex virus, lentivirus, papilloma virus, and retrovirus vectors. Some of these vectors, such as retroviral vectors, will deliver the gene therapy DNA construct to the chromosomal DNA of the patient's cells, and the gene therapy DNA construct can integrate into the chromosomal DNA; other vectors will function as episomes and the gene therapy DNA construct will remain in the cytoplasm.

Another means to increase endogenous alpha 2 polypeptide expression in a cell via gene therapy is to insert one or more enhancer elements into the alpha 2 polypeptide promoter, where the enhancer element(s) can serve to increase transcriptional activity of the alpha 2 polypeptides gene. The enhancer element(s) used will be selected based on the tissue in which one desires to activate the gene(s); enhancer elements known to confer promoter activation in that tissue will be selected. For example, if a gene encoding a

alpha 2 polypeptide is to be "turned on" in T-cells, the *lck* promoter enhancer element may be used. Here, the functional portion of the transcriptional element to be added may be inserted into a fragment of DNA containing the alpha 2 polypeptide promoter (and optionally, inserted into a vector and/or 5' and/or 3' flanking sequence(s), etc.) using standard cloning techniques. This construct, known as a "homologous recombination construct," can then be introduced into the desired cells either *ex vivo* or *in vivo*.

Gene therapy can be used to decrease alpha 2 polypeptide expression by modifying the nucleotide sequence of the endogenous promoter(s). Such modification is typically accomplished via homologous recombination methods. For example, a DNA molecule containing all or a portion of the promoter of the alpha 2 gene(s) selected for inactivation can be engineered to remove and/or replace pieces of the promoter that regulate transcription. Here, e.g., the TATA box and/or the binding site of a transcriptional activator of the promoter may be deleted using standard molecular biology techniques; such deletion can inhibit promoter activity thereby repressing transcription of the corresponding alpha 2 gene. Deletion of the TATA box or transcription activator binding site in the promoter may be accomplished by generating a DNA construct comprising all or the relevant portion of the alpha 2 polypeptide promoter(s) (from the same or a related species as the alpha 2 gene(s) to be regulated) in which one or more of the TATA box and/or transcriptional activator binding site nucleotides are mutated via substitution, deletion and/or insertion of one or more nucleotides such that the TATA box and/or activator binding site has decreased activity or is rendered completely inactive. This construct, which also will typically

contain at least about five hundred bases of DNA that correspond to the native (endogenous) 5' and 3' DNA sequences adjacent to the promoter segment that has been modified, may be introduced into the appropriate cells (either *ex vivo* or *in vivo*) either directly or via a viral vector as described above. Typically, integration of the construct into the genomic DNA of the cells will be via homologous recombination, where the 5' and 3' DNA sequences in the promoter construct can serve to help integrate the modified promoter region via hybridization to the endogenous chromosomal DNA.

Other gene therapy methods may also be employed where it is desirable to inhibit the activity of one or more alpha 2 polypeptides. For example, antisense DNA or RNA molecules, which have a sequence that is complementary to at least a portion of the selected alpha 2 polypeptide gene(s) can be introduced into the cell. Typically, each such antisense molecule will be complementary to the start site (5' end) of each selected alpha 2 gene. When the antisense molecule then hybridizes to the corresponding alpha 2 mRNA, translation of this mRNA is prevented.

Alternatively, gene therapy may be employed to create a dominant-negative inhibitor of one or more alpha 2 polypeptides. In this situation, the DNA encoding a mutant full length or truncated polypeptide of each selected alpha 2 polypeptide can be prepared and introduced into the cells of a patient using either viral or non-viral methods as described above. Each such mutant is typically designed to compete with endogenous polypeptide in its biological role.

Nucleic acid molecules of the invention may be used to map the locations of the alpha 2 gene and related genes on chromosomes. Mapping may be done by techniques known in the art, such as PCR amplification and *in situ* hybridization.

The nucleic acid molecules are also used as anti-sense inhibitors of alpha 2 expression. Such inhibition may be effected by nucleic acid molecules which are complementary to and hybridize to expression control sequences (triple helix formation) or to alpha 2 mRNA. Anti-sense probes may be designed by available techniques using the sequence of alpha 2 disclosed herein. Anti-sense inhibitors provide information relating to the decrease or absence of an alpha 2 polypeptide in a cell or organism.

Hybridization probes may be prepared using the alpha 2 sequence provided herein to screen cDNA, genomic or synthetic DNA libraries for related sequences. Regions of the DNA and/or amino acid sequence of alpha 2 that exhibit significant identity to known sequences are readily determined using sequence alignment algorithms disclosed above and those regions may be used to design probes for screening.

The nucleic acid molecules of the invention may be used for gene therapy. Nucleic acid molecules which express alpha 2 *in vivo* provide information relating to the effects of the polypeptide in cells or organisms.

Alpha 2 nucleic acid molecules, fragments, variants, and/or derivatives that do not themselves encode biologically active polypeptides may be useful as hybridization probes in diagnostic assays to test,

either qualitatively or quantitatively, for the presence of alpha 2 DNA or corresponding RNA in mammalian tissue or bodily fluid samples.

Alpha 2 polypeptide fragments, variants, and/or derivatives, whether biologically active or not, are useful for preparing antibodies that bind to an alpha 2 polypeptide. The antibodies may be used for *in vivo* and *in vitro* diagnostic purposes, including, but not limited to, use in labeled form to detect the presence of alpha 2 polypeptide in a body fluid or cell sample. The antibodies may bind to an alpha 2 polypeptide so as to diminish or block at least one activity characteristic of an alpha 2 polypeptide, or may bind to a polypeptide to increase an activity.

Description of Specific Embodiments

The invention is now described in further detail with reference to the following methods, materials, procedures and results. These examples are intended for illustration purposes only, and should not be construed as limiting the scope of the invention in any way.

Example 1
DNA Encoding Mouse and Human Alpha 2

An amino acid sequence profile for the glycoprotein hormone α polypeptide was created using the amino acid sequences of twenty different glycoprotein hormone α polypeptides from various mammals, fish and amphibians. A 6-way translation of the public EST database was searched with this profile, and one submission from mouse thymus (GenBank accession no. AA709641) was found to have a low, but potentially

significant, profile score and also to have an amino acid sequence distinct from that of the known glycoprotein hormone α polypeptides.

The bacterial stock containing this cDNA clone (no. 1224990) was obtained (Genome Systems, Inc., St. Louis, MO) and the DNA sequence (both strands) of the full insert was determined. This nucleic acid sequence (SEQ ID NO: 22) contained an open reading frame with a N-terminal predicted signal peptide followed by an amino acid sequence markedly distinct from that of the known glycoprotein hormone α polypeptides (but still maintaining the low but potentially significant glycoprotein hormone α polypeptide profile score)(SEQ ID NO: 2). There was no stop codon 5' of the methionine located at the N-terminus of the predicted signal peptide. Thus it was not possible to formally determine whether the coding region represented the full coding region of a gene, the 3' region of a larger cysteine rich protein, or non-coding region DNA (intron, 3'UTR, intergenic DNA). Repeated 5' RACE experiments using mouse thymus cDNA did not result in the identification of a 5' stop codon. A 386-base pair PCR fragment of the full insert sequence (clone 1224990) was generated using the following reaction mix and PCR conditions:

Template: ten picograms of clone 1224990 plasmid DNA.
Forward primer: 5'-ATGCCCATGGCACCACGAGTCT-3' (SEQ ID NO: 23).
Reverse primer: 5'-CTAGTAGCGGGAGAAACGGCACATATCA-3' (SEQ ID NO: 24).
Final concentration of each primer: 1.0 micromolar.
Final concentration of dNTPs: 200 micromolar.
Ten units of Taq polymerase.
Final reaction volume: 100 microliters.

Cycling conditions: 94°C for sixty seconds followed by 35 cycles of 94°C (ten seconds), 68°C (sixty seconds), and at the end of the 35th cycle an incubation at 68°C for six minutes.

The 386-base pair PCR product was then purified using agarose gel electrophoresis. Using this PCR fragment as a probe it was not possible to obtain a hybridization signal on Mouse Multiple Tissue Northern Blots even under low stringency hybridization conditions.

The computer program BLAST was used to compare the full insert DNA sequence of clone 1224990 to the DNA sequences present in the public database. This search resulted in the identification of an additional homologous mouse EST (GenBank accession no. C89076), derived from a blastocyst cDNA library, as well as two human sequences having homology to the mouse sequence. One of the human sequences was an EST from a cDNA library (GenBank accession no. AA939347), and the other consisted of 108 kilobase pairs of human genomic DNA sequence from chromosome 11 (GenBank accession no. AC000159). The homology of the mouse clone 1224990 sequence to the human genomic DNA sequence was discontinuous and resembled the typical exon-intron pattern found in genes of higher eukaryotes. Some coding region DNA sequence was assembled from this presumed human ortholog of mouse clone 1224990, and PCR was used to clone a 466-base pair fragment of this gene from human testis cDNA using the following reaction mix and PCR conditions:

Template: ten microliters of Human Testis Marathon Ready cDNA (Clontech Laboratories, Inc., Palo Alto, CA; catalog no. 7414-1).

Forward primer: 5'-GAGACATCTCCCCACTGTGTTT-3' (SEQ ID NO: 25).
Reverse primer: 5'-GTTTCCCCCAACAGAATGTCAA-3' (SEQ ID NO: 26).
Final concentration of each primer: 1.0 micromolar.
Final concentration of dNTPs: 200 micromolar.
Five units of Pfu polymerase.
Final reaction volume: 100 microliters.
Cycling conditions: 94°C for sixty seconds followed by 35 cycles of 94°C (ten seconds), 60°C (thirty seconds), 72°C (sixty seconds), and then at the end of the 35th cycle an incubation at 72°C for five minutes.

The PCR reaction was run on an agarose gel, and four distinct bands were seen. The multiple bands arose from PCR amplification of contaminating human genomic DNA present in the Human Testis Marathon Ready cDNA. The 466-base pair PCR product was isolated from the agarose gel and cloned. A clone containing sequence from the presumed human ortholog of mouse clone 1224990 was identified by sequencing. This clone was used as a template for generating a 390-base pair PCR fragment using the following reaction mix and PCR conditions:

Template: ten picograms of the testis cDNA derived PCR clone containing the 466-base pair sequence from the presumed human ortholog of mouse clone 1224990.
Forward primer: 5'-ATGCCTATGGCGTCCCCTCAAAC-3' (SEQ ID NO: 27).
Reverse primer: 5'-CTAGTAGCGAGAGAGGCGACACATGTCA-3' (SEQ ID NO: 28).
Final concentration of each primer: 1.0 micromolar.
Final concentration of dNTPs: 200 micromolar.
Ten units of Taq polymerase.
Final reaction volume: 100 microliters.

Cycling conditions: 94°C for sixty seconds, followed by 35 cycles of 94°C (ten seconds), 68°C (sixty seconds), and then at the end of the 35th cycle an incubation at 68°C for six minutes.

The 390-base pair PCR product was then purified by agarose gel electrophoresis. This 390-base pair region represented coding region sequence only (no intronic sequence or untranslated sequence). This PCR fragment was used as a probe for hybridizing to various Human Multiple Tissue northern blots (Clontech Laboratories, Inc., Palo Alto, CA). A strong signal representing a single band was obtained in the pancreas mRNA lane and a significantly weaker signal was seen in the placenta mRNA lane, under stringent hybridization conditions.

To confirm this result and determine the size of this mRNA, a Northern blot was made with human pancreas mRNA obtained from Invitrogen (Carlsbad, CA) and hybridized with the same PCR product using stringent hybridization conditions. A strong signal representing a single band was once again obtained and, from the RNA size markers that had been included in the gel, a semi-log plot of size versus migration distance gave an estimated mRNA size of 760 nucleotides. To clone the most 5' region of this human mRNA, 5' RACE was performed using the following reaction mix and PCR conditions:

Template: five microliters of Human Pancreas Marathon Ready cDNA (Clontech Laboratories, Inc., Palo Alto, CA; catalog no. 7410-1).
Forward primer: 5'-CCATCCTAATACGACTCACTATAGGGC-3' (SEQ ID NO: 29).
Reverse primer: 5'-GTTTCCCCCAACAGAATGTCAAG-3' (SEQ ID NO: 30).

Final concentration of each primer: 1.0 micromolar.
Final concentration of dNTPs: 200 micromolar.
Two and one-half units of Pfu polymerase.
Final reaction volume: fifty microliters.
Cycling conditions: 94°C for thirty seconds followed by 35 cycles of 94°C (ten seconds), 60°C (twenty seconds), 72°C (ninety seconds), and then at the end of the 35th cycle an incubation at 72°C for four minutes.

PCR products were cloned into pPCR-Script-Amp (Stratagene, La Jolla, CA) and a number of clones were sequenced. The longest of these was assembled by computer with the 3' UTR sequences present in AA939347 and AC000159, resulting in a final cDNA sequence (not including any polyA tail sequence) of 742 base pairs (SEQ ID NO: 21). Even a very short polyA tail put this sequence within the size range predicted by the Northern blot analysis, indicating that this human cDNA sequence was full length, or within a few nucleotides of being full length. From this sequence it was possible to formally define the human coding region (SEQ ID NO: 3). By amino acid homology (Figure 2) it was determined that the full mouse coding region (SEQ ID NO: 4) was contained in the full insert sequence of mouse clone 1224990. N-terminal signal peptides were predicted by computer for both the human and mouse sequences, indicating that the protein encoded by this gene was very likely to be secreted.

The following is a list and description of α2 sequences from the publicly available databases:

HUMAN:

GenBank Accession # AC000159: 108 kilobase pairs of human genomic sequence. No exons, genes or homologies are identified in this record and the full coding

region sequence of α2 is broken up by intronic sequences.

GenBank Accession # AA939347: A short EST sequence containing the 3' thirty amino acids and 3' UTR of α2. There are three clustered sequencing errors or ambiguities, resulting in the identity of one of the thirty amino acids being unassignable.

MOUSE:

GenBank Accession # AA709641: EST sequence which spans the full coding region of α2 but with sequencing errors and ambiguities the produce two frame shifts within the coding region, a missing amino acid and an unassignable amino acid.

GenBank Accession # C89076: EST sequence which spans the full coding region of α2 but with sequencing errors that produce an incorrect amino acid and an unassignable amino acid within the coding region. In addition, due to a sequencing ambiguity, the α2 stop codon is unassignable and the open reading frame extends all the way (nineteen amino acids) to the end of the EST sequence.

See also "Characterization of Gene Expression in Mouse Blastocyst Using Single-Pass Sequencing of 3995 clones" by Sasaki et al., Genomics, Volume 49, pages 167-179 (1998), which describes the cDNA library construction and subsequent sequencing effort that gave rise to the C89076 EST. The C89076 EST is not listed or reported in the Genomics paper but was found to have been entered into GenBank.

RAT:

GenBank Accession # AI111460: A short EST sequence containing the 3' forty two amino acids and 3' UTR of α2. A sequencing error converts the 3' terminal amino acid into a stop codon.

GenBank Accession # AI112649: A short EST sequence containing the 3' twenty three amino acids and 3' UTR of α2.

GenBank Accession # AI111377: A short EST sequence containing the 3' forty five amino acids and 3' UTR of α2. A sequencing error produces a frameshift near the middle of the protein coding region of the EST.

In the GenBank reports for all seven of the above entries (AC000159, AA939347, AA709641, C89076, AI111460, AI112649, AI111377), there are no predictions or identifications of any coding region, nor any homology to any known gene.

Example 2
Tissue Expression

Northern analysis was carried out to determine the expression pattern of alpha 2.

For mouse, the 386-base pair PCR product described in Example 1, above, was labeled with ^{32}P and hybridized to Clontech Mouse Multiple Tissue Northern Blots (7-day fetus, 11-day fetus, 15-day fetus, 17-day fetus, heart, brain, spleen, lung, liver, skeletal muscle, kidney and testis) using high stringency and then, separately, low stringency conditions, as follows:

High stringency:
Hybridization was for one hour at 68°C using Clontech "ExpressHyb Hybridization Solution". The blots were washed in 2x SSC, 0.1% SDS at room temperature twice for 20 minutes each time. The blots were then washed in 0.1x SSC, 0.1% SDS at 50°C for 10 minutes and exposed to film.

Low stringency:
Hybridization was for one hour at 62°C using Clontech "ExpressHyb Hybridization Solution". The blots were washed in 2x SSC, 0.1% SDS at room temperature twice for 20 minutes each time and then exposed to film.

No signal was obtained on the mouse Northerns under either high or low stringency hybridization conditions.

For human, the 390-base pair PCR product described in Example 1 was labeled with ^{32}P and hybridized, using the same high stringency conditions described above for mouse, to Clontech Human Multiple Tissue Northern Blots (pancreas, adrenal medulla, thyroid, adrenal cortex, testis, thymus, small intestine, stomach, spleen, prostate, ovary, colon, peripheral blood leucocytes, brain, heart, skeletal muscle, kidney, liver, placenta and lung) and to a Northern blot made with pituitary mRNA. A strong signal representing a single band was obtained in the pancreas mRNA lane and the pituitary mRNA lane. A significantly weaker signal was seen in the placenta mRNA lane.

In situ hybridization was done to further determine sites of α2 gene expression. A panel of normal embryonic (E10.5 through E18.5) and adult mouse tissues and adult rhesus monkey tissues were fixed in

4% paraformaldehyde, embedded in paraffin, and sectioned at 5 micrometers. Prior to *in situ* hybridization, tissues were permeabilized with 0.2M HCL, followed by digestion with Proteinase K and acetylation with triethanolamine and acetic anhydride. Sections were hybridized overnight at 55°C with a ^{33}P-labeled antisense RNA probe complementary to either the mouse or human (for rhesus tissues) α2 sequence and with sense (control) probes. The antisense and sense ^{33}P-labeled RNA probes were obtained by *in vitro* transcription of plasmid DNAs containing either the mouse α2 cDNA or the human α2 cDNA.

Following hybridization, sections were washed in buffer, treated with RNaseA to remove unhybridized probe, then subjected to a high stringency wash in 0.1X SSC at 55°C. Slides were dipped in Kodak NTB2 emulsion, exposed at 4°C for two-three weeks, developed, and then counterstained. Sections were examined with darkfield and standard illumination to allow simultaneous evaluation of tissue morphology and hybridization signal. The following tissues were then examined:

Mouse tissues: Brain (1 sagittal, 2 coronal sections); GI tract(esophagus, stomach, duodenum, jejunum, ileum, proximal & distal colon); pituitary; liver; lung; heart; spleen; thymus; lymph nodes; kidney; adrenal; bladder; pancreas; salivary gland; male and female reproductive organs (ovary, oviduct and uterus in the female; testis, epididymus, prostate, seminal vesicle and vas deferens in the male); BAT & WAT (subcutaneous, peri-renal); bone (femur); skin; breast; and skeletal muscle.

Rhesus tissues: adrenal gland; liver; gall bladder; intestine; pancreas; and salivary gland.

Both mouse and human antisense probes produced positive signal detectable above a very low level of background seen with the sense strand controls. In the embryonic mouse, no signal was observed in any major organs from E8.5 through E18.5. At E15.5 and E18.5, signal was present over scattered cells adjacent to some of the developing bones of the head and teeth. In the adult mouse, a moderate level of signal was present in the adrenal cortex. A lower level of signal was detectable in the anterior and intermediate lobes of the pituitary as well as in intestinal epithelium at the level of the crypts. In addition, grain density was slightly above background in developing sperm within the seminiferous tubules of the testis and in granulosa cells surrounding developing follicles in the ovary.

In rhesus tissues, moderate signal was noted in the adrenal cortex, gall bladder epithelium, and in the intestinal epithelium primarily at the level of the crypts.

Example 3

Production of Alpha 2 Polypeptides

A. Bacterial Expression

PCR is used to amplify template DNA sequences encoding an alpha 2 polypeptide using primers corresponding to the 5' and 3' ends of the sequence. The amplified DNA products may be modified to contain restriction enzyme sites to allow for insertion into expression vectors. PCR products are gel purified and inserted into expression vectors using standard recombinant DNA methodology. An exemplary vector, such as pAMG21 (ATCC No. 98113)

containing the lux promoter and a gene encoding kanamycin resistance, is digested with BamHI and NdeI for directional cloning of inserted DNA. The ligated mixture is transformed into an *E. coli* host strain by electroporation and transformants are selected for kanamycin resistance. Plasmid DNA from selected colonies is isolated and subjected to DNA sequencing to confirm the presence of the insert.

Transformed host cells are incubated in 2XYT medium containing 30 µg/ml kanamycin at 30°C prior to induction. Gene expression is induced by addition of N-(3-oxohexanoyl)-dl-homoserine lactone to a final concentration of 30 ng/ml followed by incubation at either 30°C or 37°C for six hours. Expression of alpha 2 polypeptide is evaluated by centrifugation of the culture, resuspension and lysis of the bacterial pellets, and analysis of host cell proteins by SDS-polyacrylamide gel electrophoresis.

Inclusion bodies containing alpha 2 polypeptide are purified as follows:

Bacterial cells are pelleted by centrifugation and resuspended in water. The cell suspension is lysed by sonication and pelleted by centrifugation at 195,000 x g for 5 to 10 minutes. The supernatant is discarded and the pellet washed and transferred to a homogenizer. The pellet is homogenized in 5 ml. of a Percoll solution (75% liquid Percoll. 0.15M NaCl) until uniformly suspended and then diluted and centrifuged at 21,600 x g for 30 minutes. Gradient fractions containing the inclusion bodies are recovered and pooled. The isolated inclusion bodies are analyzed by SDS-PAGE.

A single band on an SDS polyacrylamide gel corresponding to *E. coli*-produced alpha 2 polypeptide is excised from the gel and N-terminal amino acid sequence is determined essentially as described by Matsudaira et al., J. Biol. Chem. 262, 10-35 (1987).

B. Mammalian Cell Production

PCR is used to amplify template DNA sequences encoding an alpha 2 polypeptide using primers corresponding to the 5' and 3'ends of the sequence. The amplified DNA products may be modified to contain restriction enzyme sites to allow for insertion into expression vectors. PCR products are gel purified and inserted into expression vectors using standard recombinant DNA methodology. An exemplary expression vector, pCEP4 (Invitrogen), which contains an Epstein-Barr virus origin of replication, may be used for expression of alpha 2 in 293-EBNA-1 cells. Amplified and gel purified PCR products are ligated into pCEP4 vector and lipofected into 293-EBNA cells. The transfected cells are selected in 100 microgram/ml hygromycin and the resulting drug-resistant cultures are grown to confluence. The cells are then cultured in serum-free media for 72 hours and the conditioned media removed and alpha 2 protein expression analyzed by SDS-PAGE.

Alpha 2 protein expression may be detected by silver staining. Alternatively, alpha 2 is produced as a fusion protein with an epitope tag, such as an IgG constant domain or a FLAG epitope, which may be detected by Western blot analysis using antibodies to the tag peptide.

Alpha 2 polypeptides may be excised from an SDS-polyacrylamide gel or alpha 2 fusion proteins are

purified by affinity chromatography to the epitope tag, and then subjected to N-terminal amino acid sequence analysis as described above.

Example 4
Production of Anti-alpha 2 Antibodies

Antibodies to alpha 2 polypeptides may be obtained by immunization with purified protein or with alpha 2 peptides produced by biological or chemical synthesis. Procedures for generating antibodies can be those described in Hudson and Hay, "Practical Immunology, Second Edition", Blackwell Scientific Publications (1980).

Animals (typically mice or rabbits) are injected with an alpha 2 antigen and those with sufficient serum titer levels as determined by enzyme-linked immunosorbent assays (EIA) are selected for hybridoma production. Spleens of immunized animals are collected and prepared as single cell suspensions from which splenocytes are recovered. The splenocytes are fused to mouse myeloma cells (such as Sp2/0-Ag14 cells), allowed to incubate in Dulbeccos' Modified Eagle' Medium (DMEM) with 200 U/ml of penicillin, 200 microgram/ml of streptomycin sulfate, and 4 mM of glutamine, then incubated in HAT selection medium (hypoxanthine; aminopterin; thymidine). After selection, tissue culture supernatants are taken from each fusion well and tested for alpha 2 antibody production by EIA.

Alternative procedures for obtaining anti-alpha 2 antibodies may also be employed, such as immunization of transgenic mice harboring human Ig loci for production of fully human antibodies, and screening of synthetic antibody libraries, such as

those generated by mutagenesis of an antibody variable domain.

Example 5

In Vivo Functional Test of Signal Peptide

A signal peptide test was done to determine whether the predicted α2 signal peptide could function as such *in vivo*. The general procedure was to replace the signal peptide of human CD2 (a known plasma membrane transmembrane protein) with the putative signal peptide of mouse α2, and then to determine whether this fusion protein was translocated to the plasma membrane in mammalian cells. Detection of CD2 on the surface of mammalian cells was done using a CD2-specific antibody. The protocol was as follows:

The six oligonucleotides listed below, representing the sequence of the mouse α2 putative signal peptide and some 3' flanking region, were synthesized, kinased, annealed together, and then cloned into a vector containing a human CD2 gene lacking its native signal peptide.

SEQ ID NO: 31:
5'-AATTCCACCATGCCCATGGCACCACGAGTCTTGCTCCTTTGCCTGCTGG-3'

SEQ ID NO: 32:
5'-GCCTGGCAGTCACTGAAGGGCATAGCCCAGAGACAGCCAT-3'

SEQ ID NO: 33:
5'-CCCAGGCTGCCACTTGCACCCCTTCT-3'

SEQ ID NO: 34:
5'-CTAGAGAAGGGGTGCAAGTGGCAGCCTGGGATGGCTGTCTCTGGG-3'

SEQ ID NO: 35:
5'-CTATGCCCTTCAGTGACTGCCAGGCCCAGCAGGCAAAGGA-3'

SEQ ID NO: 36:
5'-GCAAGACTCGTGGTGCCATGGGCATGGTGG-3'

A number of clones were sequenced, and a clone with the mouse putative α2 signal peptide region fused to truncated CD2 was identified. The amino acid sequence (SEQ ID NO: 37) of the fusion protein encoded by this clone (mouse-α2-human-CD2) is shown in Figure 6. 293T cells were transfected with mouse-α2-human-CD2 DNA using Lipofectamine (Life Technologies, Gaithersburg, MD). After three days of growth, the cells were washed and then stained with an FITC-labeled anti-human-CD2 antibody (Becton Dickinson Immunocytometry Systems, San Jose, CA). FITC fluorescence was then determined using an ACAS Ultima Fluorescent Confocal Microscope (Meridian Instruments, Inc., Okemos, MI).

Based on the intense fluorescence of 293T cells that had been transfected with the mouse-α2-human-CD2 DNA (see Figure 7C), it was concluded that the predicted α2 signal peptide functions as an efficient signal peptide *in vivo*.

While the present invention has been described in terms of preferred embodiments, it is understood that variations and modifications thereof will occur to those skilled in the art. Therefore, it is intended that the appended claims cover all such equivalent variations which come within the scope of the invention as claimed.

CLAIMS

WHAT IS CLAIMED IS:

1. An isolated alpha 2 polypeptide comprising the amino acid sequence selected from the group consisting of:

a) the amino acid sequence set forth in SEQ ID NO: 1, 5, 6, 7, 8, 9, 10, 11 or 12, optionally further comprising an amino-terminal methionine;

b) a fragment of SEQ ID NO: 1, 5, 6, 7, 8, 9, 10, 11 or 12 comprising about fifty or more amino acid residues, wherein the fragment has an activity of the polypeptides of SEQ ID NOS: 1 and 5-12;

c) an ortholog of SEQ ID NO: 1, 5, 6, 7, 8, 9, 10, 11 or 12; and

d) an analog of (a), (b) or (c) which is at least eighty percent identical at the amino acid level.

2. The isolated polypeptide of claim 1 in which the analog of (a), (b) or (c) is at least ninety percent identical at the amino acid level.

3. The isolated polypeptide of claim 1 in which the analog of (a), (b) or (c) is at least ninety five percent identical at the amino acid level.

4. An isolated polypeptide according to claim 1 which is an ortholog of the polypeptide of SEQ ID NO: 1 and has the amino acid sequence of SEQ ID NO: 2.

5. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

a) a nucleotide sequence encoding the polypeptide of SEQ ID NO: 1, 2, 5, 6, 7, 8, 9, 10, 11 or 12;

b) the nucleotide sequence of SEQ ID NO: 3, 4, 13, 14, 15, 16, 17, 18, 19, 20 or 21;

c) a nucleotide sequence encoding a polypeptide that is at least eighty percent identical to the polypeptide of SEQ ID NO: 1, 5, 6, 7, 8, 9, 10, 11 or 12 wherein the polypeptide has an activity of the polypeptide of SEQ ID NOS: 1 and 5-12;

d) a nucleotide sequence of (b) or (c) encoding a polypeptide fragment of about fifty or more amino acid residues wherein the polypeptide fragment has an activity of the polypeptide of SEQ ID NOS: 1 and 5-12;

e) a nucleotide sequence which hybridizes under stringent conditions to the complement of any of (a) -(d); and

f) a nucleotide sequence complementary to any of (a), (b) or (c).

6. The isolated nucleic acid of claim 5 in which (d) encodes a polypeptide that is at least ninety percent identical to the polypeptide of SEQ ID NO: 1, 5, 6, 7, 8, 9, 10, 11 or 12.

7. The isolated nucleic acid of claim 6 in which (d) encodes a polypeptide that is at least

ninety five percent identical to the polypeptide of SEQ ID NO: 1, 5, 6, 7, 8, 9, 10, 11 or 12.

8. An expression vector comprising the nucleic acid molecule of Claim 5.

9. A host cell comprising the expression vector of Claim 8.

10. The host cell of Claim 9 which is a eukaryotic cell.

11. The host cell of Claim 9 which is a prokaryotic cell.

12. A process for producing a protein comprising growing a culture of the host cell of Claim 9 in suitable culture medium and isolating the protein from the culture.

13. A process for producing the polypeptide of claim 1 by genetically engineering a mammalian cell to express the polypeptide from one or more of the cell's endogenous genes.

14. The process of claim 13 which involves the introduction of a promoter and/or enhancer to cause or increase the expression of the polypeptide.

15. The process of claim 13 which involves the introduction of one or more genes encoding one or more naturally occurring or non-naturally occurring polypeptides that cause or increase expression of the cell's endogenous alpha 2 gene.

16. The process of claim 13 which involves the introduction of one or more recombination

sequences such that recombination between such sequences results in an insertion, deletion, inversion or translocation event that causes or increases expression of the cell's endogenous alpha 2 gene.

17. A polypeptide produced by the process of Claim 12, 13, 14, 15 or 16.

18. An antibody or fragment thereof that specifically binds the polypeptide of Claim 1.

19. The antibody of Claim 18 that is a monoclonal antibody.

20. A binding peptide or fragment thereof that specifically binds the polypeptide of claim 1.

21. A polypeptide derivative of the polypeptide of Claim 1.

22. The polypeptide derivative of Claim 21 in which the polypeptide is covalently modified with a water-soluble polymer.

23. The polypeptide of Claim 22 wherein the water-soluble polymer is selected from the group consisting of polyethylene glycol and dextran.

24. The polypeptide derivative of Claim 21 in which the glycosylation has been modified.

25. A fusion polypeptide comprising the polypeptide of Claim 1 fused to a heterologous amino acid sequence.

26. The fusion polypeptide of Claim 25 wherein the heterologous amino acid sequence is an IgG constant domain or fragment thereof.

27. A composition comprising the polypeptide of Claim 1, 21, 24 or 25 and a pharmaceutically acceptable formulation agent.

28. The composition of Claim 27 wherein the polypeptide comprises the amino acid sequence of SEQ ID NO: 1, 5, 6, 7, 8, 9, 10, 11 or 12.

29. A method for treating, preventing or ameliorating a medical condition comprising administering to a patient a polypeptide according to Claim 1, 21, 24 or 25, or an antibody or fragment thereof according to claim 18 or 19, or a binding peptide or fragment thereof according to claim 20.

30. A method according to claim 29 in which the medical condition is selected from the group consisting of obesity, wasting syndromes, cachexia, gastrointestinal disorders, ulcers, diabetes, hypertension, immune system dysfunction, excessive inflammation, susceptibility to infection, AIDS, poor wound healing, asthma, arthritis, allergies, shock, tissue damage/degeneration, cancers, hyperplasias/hypertrophies, infertility, fertility, contraception and impotence.

31. A method for treating, preventing or ameliorating a medical condition comprising administering to a patient a nucleic acid according to claim 3.

32. A method according to claim 31 in which the medical condition is selected from the group

consisting of obesity, wasting syndromes, cachexia, gastrointestinal disorders, ulcers, diabetes, hypertension, immune system dysfunction, excessive inflammation, susceptibility to infection, AIDS, poor wound healing, asthma, arthritis, allergies, shock, tissue damage/degeneration, cancers, hyperplasias/hypertrophies, infertility, fertility, contraception and impotence.

33. A method for treating, preventing or ameliorating a medical condition comprising treating cells, tissues or organs *ex vivo* with the polypeptide of claim 1, 21 or 25 and implanting the treated cells, tissues or organs into the patient being treated.

34. A method according to claim 33 in which the medical condition is selected from the group consisting of obesity, wasting syndromes, cachexia, gastrointestinal disorders, ulcers, diabetes, hypertension, immune system dysfunction, excessive inflammation, susceptibility to infection, AIDS, poor wound healing, asthma, arthritis, allergies, shock, tissue damage/degeneration, cancers, hyperplasias/hypertrophies, infertility, fertility, contraception and impotence.

35. A process for treating, preventing or ameliorating a medical condition in a patient by genetically engineering a human cell or cells to express the polypeptide of claim 1, 21 or 25 and implanting the cell or cells into the patient.

36. A method according to claim 35 in which the medical condition is selected from the group consisting of obesity, wasting syndromes, cachexia, gastrointestinal disorders, ulcers, diabetes, hypertension, immune system dysfunction, excessive

inflammation, susceptibility to infection, AIDS, poor wound healing, asthma, arthritis, allergies, shock, tissue damage/degeneration, cancers, hyperplasias/hypertrophies, infertility, fertility, contraception and impotence.

37. A method of diagnosing a medical condition or a susceptibility to a medical condition in a subject comprising:

a) determining the presence or amount of expression of the polypeptide of claim 1 or of the nucleic acid of claim 3; and

b) diagnosing a medical condition or a susceptibility to a medical condition based on the presence or amount of expression of the polypeptide or nucleic acid.

38. The method of claim 37 which involves the use of RT-PCR or nucleic acid hybridization to measure the presence and amount of said nucleic acid.

39. The method of claim 37 which involves the use of an antibody or binding peptide specific to the polypeptide to measure the presence and amount of said polypeptide.

40. A method according to claim 37 in which the medical condition is selected from the group consisting of obesity, wasting syndromes, cachexia, gastrointestinal disorders, ulcers, diabetes, hypertension, immune system dysfunction, excessive inflammation, susceptibility to infection, AIDS, poor wound healing, asthma, arthritis, allergies, shock, tissue damage/degeneration, cancers,

hyperplasias/hypertrophies, infertility, fertility, contraception and impotence.

41. A method of identifying a compound which binds to a polypeptide comprising:

a) contacting the polypeptide of Claim 1 with a compound; and

b) determining the extent of binding of the polypeptide to the compound.

42. The method of Claim 41 further comprising determining the activity of the polypeptide when bound to the compound.

43. A method of identifying a receptor for the polypeptide of claim 1 comprising:

a) contacting the polypeptide with various cell types;

b) determining the extent of binding of the polypeptide to cell types; and

c) identifying the receptor on the surface of the cell types responsible for binding the polypeptide.

44. A method of modulating levels of a polypeptide in a mammal comprising administering to the mammal the nucleic acid molecule of Claim 3.

45. A non-human mammal which has been genetically engineered to specifically alter the level of expression of the polypeptide of claim 1.

46. A method for determining whether a compound increases or decreases alpha 2 polypeptide activity in a mammal comprising exposing the mammal to the compound, and measuring alpha 2 polypeptide activity in said mammal.

FIG. 1

Human glycoprotein hormone α2 polypeptide:

MPMASPQTLVLYLLVLAVTEAWGQEAVIPGCHLHPFNVTVRSDRQGTCQGSHVAQACVGHCESSA

FPSRYSVLVASGYRHNITSVSQCCTISGLKKVKVQLQCVGSRREELEIFTARACQCDMCRLSRY

Mouse glycoprotein hormone α2 polypeptide:

MPMAPRVLLLCLLGLAVTEGHSPETAIPGCHLHPFNVTVRSDRLGTCQGSHVAQACVGHCESSA

FPSRYSVLVASGYRHNITSSSQCCTISSLRKVRVWLQCVGNQRGELEIFTARACQCDMCRFSRY

FIG. 2

GAP OF: Human α2 CHECK: 2695 FROM: 1 TO: 129

TO: Mouse α2 CHECK: 2533 FROM: 1 TO: 128

SYMBOL COMPARISON TABLE:
/GCGDISK/GCG9/GCGCORE/DATA/RUNDATA/BLOSUM62.CMP
COMPCHECK: 6430

GAP WEIGHT:	12	AVERAGE MATCH:	2.912
LENGTH WEIGHT:	4	AVERAGE MISMATCH:	-2.003
QUALITY:	552	LENGTH:	129
RATIO:	4.313	GAPS:	1
PERCENT SIMILARITY:	85.156	**PERCENT IDENTITY:**	**83.594**

MATCH DISPLAY THRESHOLDS FOR THE ALIGNMENT(S):
| = IDENTITY
: = 2
. = 1

Human α2 x Mouse α2

```
   1  MPMASPQTLVLYLLVLAVTEAWGQEAVIPGCHLHPFNVTVRSDRQGTCQG   50
      ||||  |.  |.|  ||  |||||     |  ||||||||||||||||  |||||
   1  MPMA.PRVLLLCLLGLAVTEGHSPETAIPGCHLHPFNVTVRSDRLGTCQG   49

  51  SHVAQACVGHCESSAFPSRYSVLVASGYRHNITSVSQCCTISGLKKVKVQ   100
      |||||||||||||||||||||||||||||||||||||  |||||||  |:||:|
  50  SHVAQACVGHCESSAFPSRYSVLVASGYRHNITSSSQCCTISSLRKVRVW   99

 101  LQCVGSRREELEIFTARACQCDMCRLSRY   129
      |||||..|  ||||||||||||||||  |||
 100  LQCVGNQRGELEIFTARACQCDMCRFSRY   128
```

FIG. 3A

```
GAP of:   human α     check: 7064   from: 1   to: 92
    to:   human α2    check: 1100   from: 1   to: 106

symbol comparison table:
/gcgdisk/gcg9/gcgcore/data/rundata/blosum62.cmp
compcheck: 6430

        gap weight:      9      average match:  2.912
     length weight:      2   average mismatch: -2.003

           quality:     53              length:    106
             ratio:  0.576               gaps:      4
percent similarity: 33.696   percent identity: 27.174

       match display thresholds for the alignment(s):
                  | = identity
                  : =   2
                  . =   1

human α   x   human α2


   1   ....APDVQDCP.ECTLQENPFFSQPGAPILQ.CMGCCFSRAYPTPLRSK   44
           |     |   |.. .   .  |. : | |.| | | |:|.
   1   QEAVIPGCHLHPFNVTVRSDRQGTCQGSHVAQACVGHCESSAFPSRYSVL   50

  45   KTMLVQKNVTSESTCCVAKSYNRVTV....MGGFKVE....NHTACHCST   86
         . |:|| | ||       :| |    .|  : |        || |
  51   VASGYRHNITSVSQCCTISGLKKVKVQLQCVGSRREELEIFTARACQCDM   100

  87   CYYHKS  92
       |   :
 101   CRLSRY  106
```

FIG. 3B

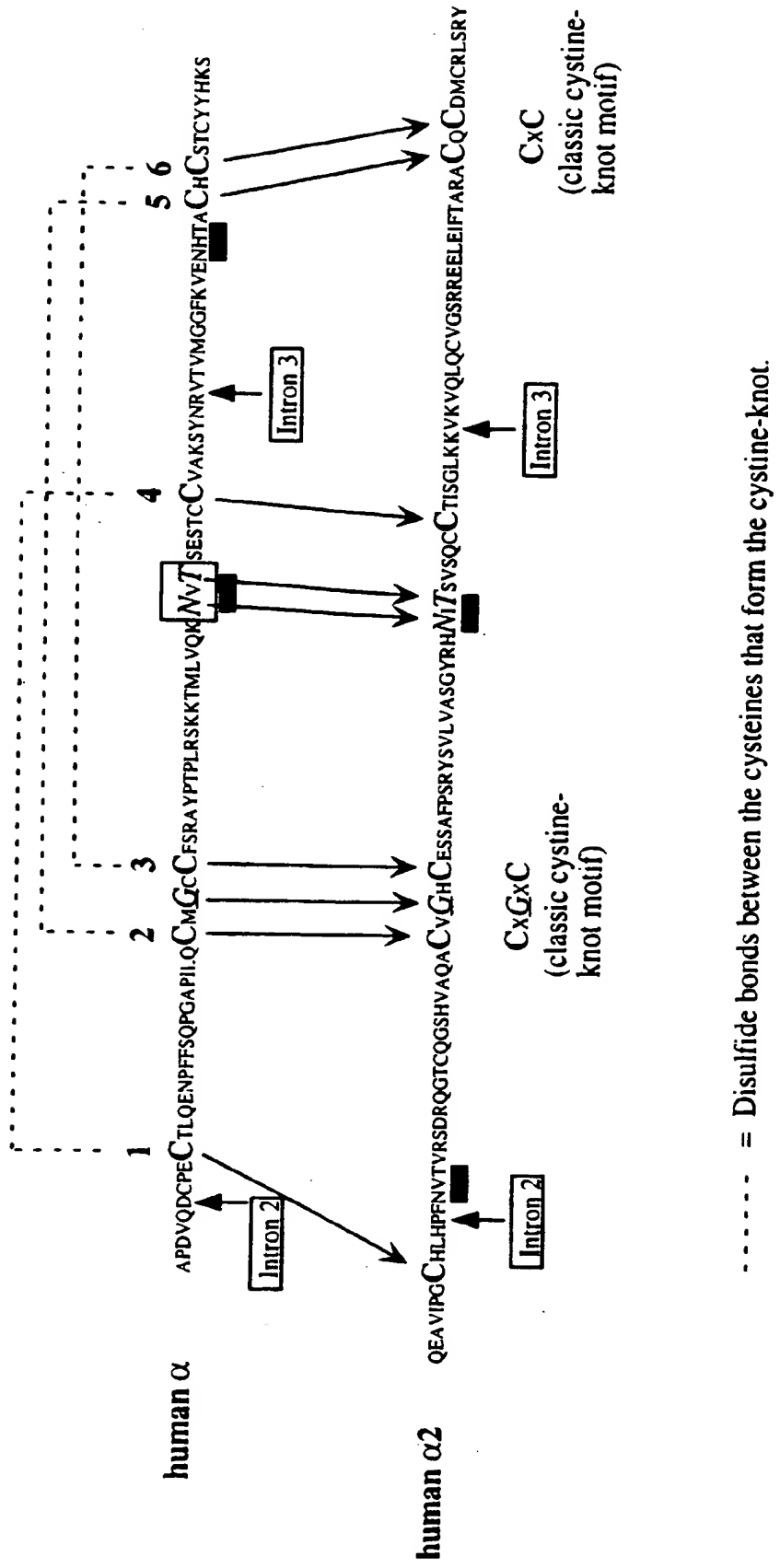

human α
human α2
1
2
3
4
5 6
APDVQDCPECTLQENPFFSQPGAPILQCMGCCFSRAYPTPLRSKKTMLVQKNVTSESTCCVAKSYNRVTVMGGFKVENHTACHCSTCYYHKS
QEAVIPGCHLHPFNVTVRSDRQGTCQGSHVAQACVGHCESSAFPSRYSVLVASGYRHNVTSVSQCCTISGLKKVKVQLQCVGSRREELEIFTARACQCDMCRLSRY
Intron 2
Intron 2
Intron 3
Intron 3
CXGXC
(classic cystine-knot motif)
CXC
(classic cystine-knot motif)
----- = Disulfide bonds between the cysteines that form the cystine-knot.
■ = asparagine (N) glycosylation NxT motifs.

FIG. 4

Human α2 cDNA

1 gcaggaggcacaggaaaactgcaagccgcttctgttcctgggcctcggaagtg (*intron 1*) **ATG** cctatggcgtcccctcaaaccctggtcctctatctgctggtcctg

101 gcagtcactgaagcctgggccaggaggcagtcatcccaggctgccacttgcacc (*intron 2*) acttcaatgtgacagtgcgaagtgaccgccaaggcacctgccagg

201 gctcccacgtggcacaggcctgtgtgggccactgtgagtccagcgcctccattctcggtactatgtgctggtggccagtggttaccgacacaacatcac

301 ctccgtctctcagtgctgtgaccatcagtggcctgaagaag (*intron 3*) gtcaaagtacagctgcagtgtgtgggggagccggagggaggagctcgagatcttcacggcc

401 agggcctgccagtgtgacatgtgtcgcctctctctgctac **TAG** ccatcctctcctcccctcctcctcccctgggtcacagggcttgacattctggtgggggA

501 aacctgttcaagattcaaaaactggaaggagtcccagccctgatggttacttgctatggaatttttttaaataagggagggttgttccagctttgat

601 cctttgtaagatttttgtgactgtcacctgagaagaggggagtttctgcttctttccctgccttgcctggcacttctaaccaatctttcatcatttttact

701 tccctctttgcccttaccctaaataaagcaagcagttcttg (poly A tail)

FIG. 5

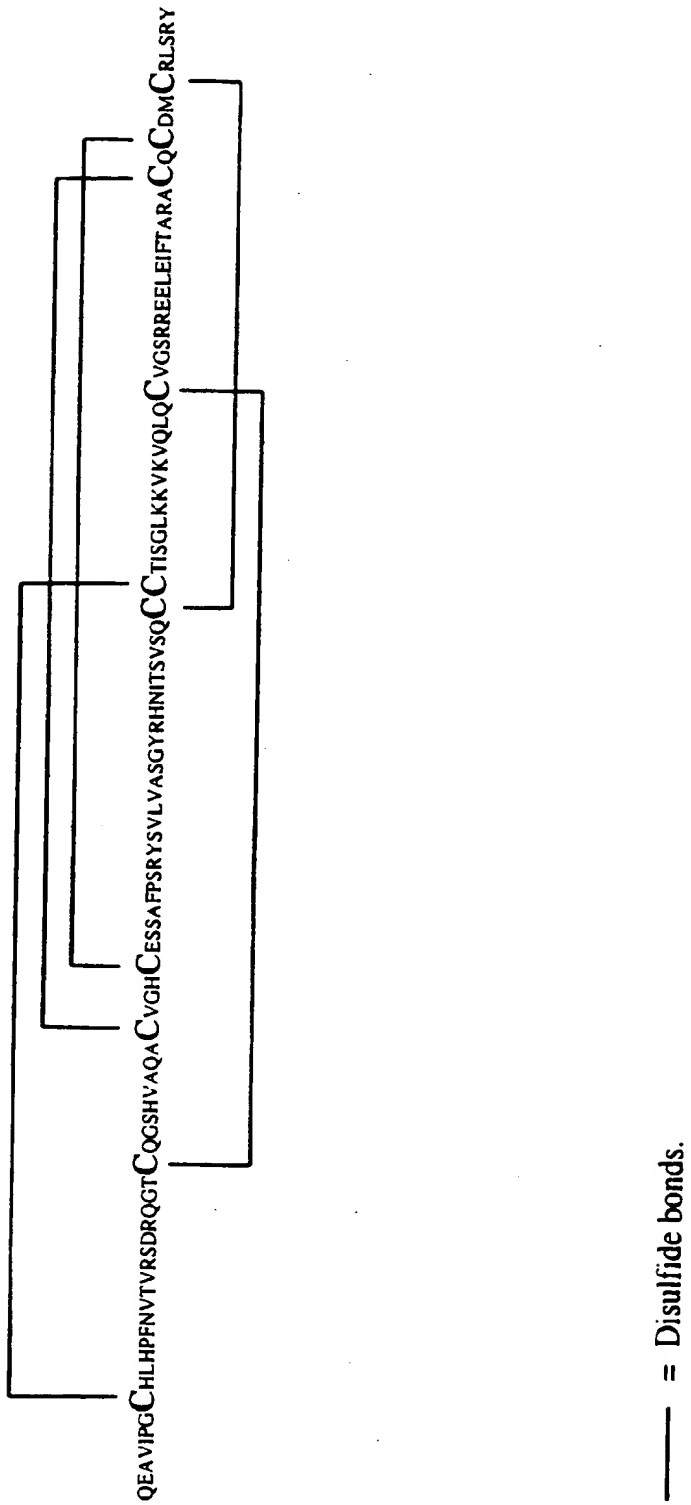

human α2
QEAVIPGCHLHPFNVTVRSDRQGTCQGSHVAQACVGHCESSAFPSRYSVLVASGYRHNITSVSQCCTISGLKKVKVQLQCVGSRREELEIFTARACQCDMCRLSRY
= Disulfide bonds.

FIG. 6

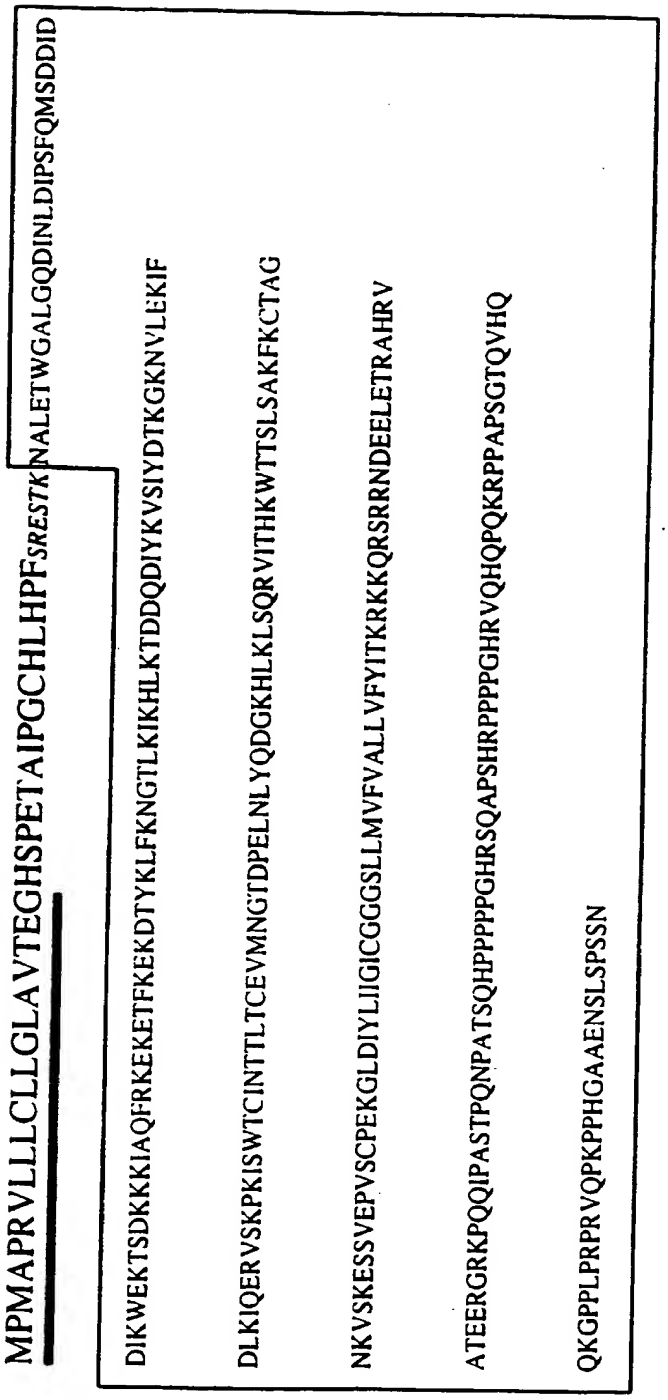


FIG. 7A

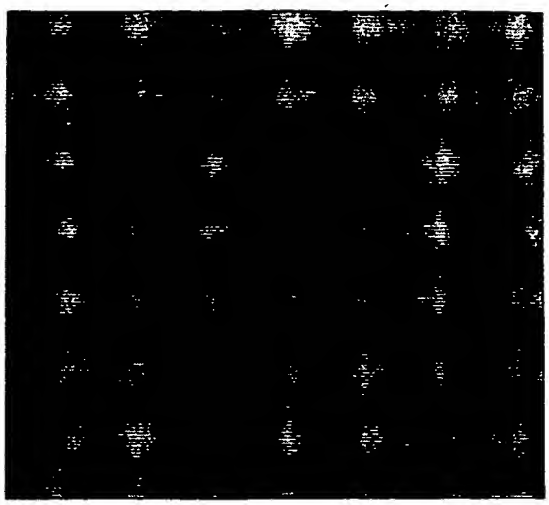


FIG. 7B

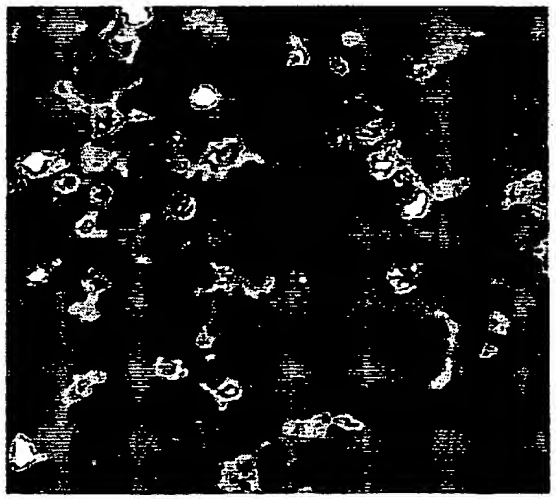
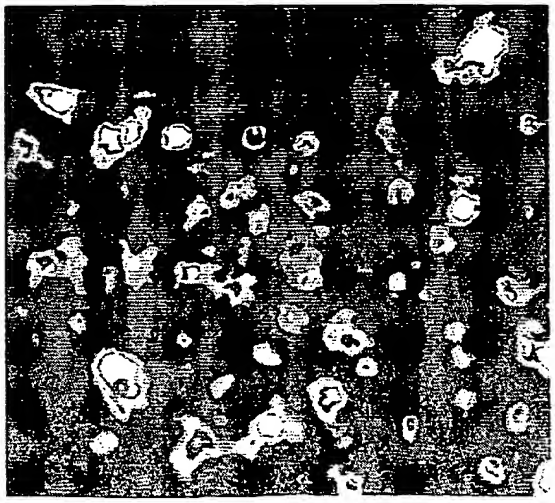


FIG. 7C



SEQUENCE LISTING

<110> Paszty, Christopher J. R.
Bass, Michael B.
Cao, Jin
Luethy, Roland
Samal, Babru B.
Schultz, Henry J.
Scully, Sheila
Welcher, Andrew

<120> GLYCOPROTEIN HORMONE FAMILY MEMBER

<130> A-589

<140>
<141>

<160> 37

<170> PatentIn Ver. 2.0

<210> 1
<211> 129
<212> PRT
<213> Human

<400> 1

Met	Pro	Met	Ala	Ser	Pro	Gln	Thr	Leu	Val	Leu	Tyr	Leu	Leu	Val	Leu
1				5					10					15	
Ala	Val	Thr	Glu	Ala	Trp	Gly	Gln	Glu	Ala	Val	Ile	Pro	Gly	Cys	His
			20					25					30		
Leu	His	Pro	Phe	Asn	Val	Thr	Val	Arg	Ser	Asp	Arg	Gln	Gly	Thr	Cys
		35					40					45			
Gln	Gly	Ser	His	Val	Ala	Gln	Ala	Cys	Val	Gly	His	Cys	Glu	Ser	Ser
	50					55					60				
Ala	Phe	Pro	Ser	Arg	Tyr	Ser	Val	Leu	Val	Ala	Ser	Gly	Tyr	Arg	His
65					70					75					80
Asn	Ile	Thr	Ser	Val	Ser	Gln	Cys	Cys	Thr	Ile	Ser	Gly	Leu	Lys	Lys
				85					90					95	
Val	Lys	Val	Gln	Leu	Gln	Cys	Val	Gly	Ser	Arg	Arg	Glu	Glu	Leu	Glu
			100					105					110		
Ile	Phe	Thr	Ala	Arg	Ala	Cys	Gln	Cys	Asp	Met	Cys	Arg	Leu	Ser	Arg
		115					120					125			

Tyr

<210> 2
<211> 128
<212> PRT
<213> Mouse

<400> 2
Met Pro Met Ala Pro Arg Val Leu Leu Leu Cys Leu Leu Gly Leu Ala
1 5 10 15

Val Thr Glu Gly His Ser Pro Glu Thr Ala Ile Pro Gly Cys His Leu
20 25 30

His Pro Phe Asn Val Thr Val Arg Ser Asp Arg Leu Gly Thr Cys Gln
35 40 45

Gly Ser His Val Ala Gln Ala Cys Val Gly His Cys Glu Ser Ser Ala
50 55 60

Phe Pro Ser Arg Tyr Ser Val Leu Val Ala Ser Gly Tyr Arg His Asn
65 70 75 80

Ile Thr Ser Ser Ser Gln Cys Cys Thr Ile Ser Ser Leu Arg Lys Val
85 90 95

Arg Val Trp Leu Gln Cys Val Gly Asn Gln Arg Gly Glu Leu Glu Ile
100 105 110

Phe Thr Ala Arg Ala Cys Gln Cys Asp Met Cys Arg Phe Ser Arg Tyr
115 120 125

<210> 3
<211> 387
<212> DNA
<213> Human

<400> 3
atgcctatgg cgtcccctca aaccctggtc ctctatctgc tggtcctggc agtcactgaa 60
gcctggggcc aggaggcagt catcccaggc tgccacttgc acccccttcaa tgtgacagtg 120
cgaagtgacc gccaaggcac ctgccagggc tcccacgtgg cacaggcctg tgtgggccac 180
tgtgagtcca gcgccttccc ttctcggtac tctgtgctgg tggccagtgg ttaccgacac 240
aacatcacct ccgtctctca gtgctgcacc atcagtggcc tgaagaaggt caaagtacag 300
ctgcagtgtg tggggagccg gagggaggag ctcgagatct tcacggccag ggcctgccag 360
tgtgacatgt gtcgcctctc tcgctac 387

<210> 4
<211> 384
<212> DNA
<213> Mouse

<400> 4
atgcccatgg caccacgagt cttgctcctt tgcctgctgg gcctggcagt cactgaaggg 60
catagcccag agacagccat cccaggctgc cacttgcacc ccttcaatgt gacggtgcgc 120

agtgatcgcc tcggcacttg ccagggctcc cacgtggcac aggcctgtgt aggacactgt 180
gagtctagtg ctttcccttc ccggtactct gtgctggtgg ccagtggcta tcggcacaac 240
atcacctctt cctcccagtg ctgcaccatc agcagcctca gaaaggtgag ggtgtggctg 300
cagtgcgtgg ggaaccagcg tggggagctt gagatcttta ctgcaagggc ctgccagtgt 360
gatatgtgcc gtttctcccg ctac 384

<210> 5
<211> 109
<212> PRT
<213> Human

<400> 5
Ala Trp Gly Gln Glu Ala Val Ile Pro Gly Cys His Leu His Pro Phe
1 5 10 15

Asn Val Thr Val Arg Ser Asp Arg Gln Gly Thr Cys Gln Gly Ser His
20 25 30

Val Ala Gln Ala Cys Val Gly His Cys Glu Ser Ser Ala Phe Pro Ser
35 40 45

Arg Tyr Ser Val Leu Val Ala Ser Gly Tyr Arg His Asn Ile Thr Ser
50 55 60

Val Ser Gln Cys Cys Thr Ile Ser Gly Leu Lys Lys Val Lys Val Gln
65 70 75 80

Leu Gln Cys Val Gly Ser Arg Arg Glu Glu Leu Glu Ile Phe Thr Ala
85 90 95

Arg Ala Cys Gln Cys Asp Met Cys Arg Leu Ser Arg Tyr
100 105

<210> 6
<211> 108
<212> PRT
<213> Human

<400> 6
Trp Gly Gln Glu Ala Val Ile Pro Gly Cys His Leu His Pro Phe Asn
1 5 10 15

Val Thr Val Arg Ser Asp Arg Gln Gly Thr Cys Gln Gly Ser His Val
20 25 30

Ala Gln Ala Cys Val Gly His Cys Glu Ser Ser Ala Phe Pro Ser Arg
35 40 45

Tyr Ser Val Leu Val Ala Ser Gly Tyr Arg His Asn Ile Thr Ser Val
50 55 60

Ser Gln Cys Cys Thr Ile Ser Gly Leu Lys Lys Val Lys Val Gln Leu
65 70 75 80

```
Gln Cys Val Gly Ser Arg Arg Glu Glu Leu Glu Ile Phe Thr Ala Arg
                85                  90                  95

Ala Cys Gln Cys Asp Met Cys Arg Leu Ser Arg Tyr
            100                 105


<210> 7
<211> 107
<212> PRT
<213> Human

<400> 7
Gly Gln Glu Ala Val Ile Pro Gly Cys His Leu His Pro Phe Asn Val
  1               5                  10                  15

Thr Val Arg Ser Asp Arg Gln Gly Thr Cys Gln Gly Ser His Val Ala
            20                  25                  30

Gln Ala Cys Val Gly His Cys Glu Ser Ser Ala Phe Pro Ser Arg Tyr
         35                  40                  45

Ser Val Leu Val Ala Ser Gly Tyr Arg His Asn Ile Thr Ser Val Ser
     50                  55                  60

Gln Cys Cys Thr Ile Ser Gly Leu Lys Lys Val Lys Val Gln Leu Gln
 65                  70                  75                  80

Cys Val Gly Ser Arg Arg Glu Glu Leu Glu Ile Phe Thr Ala Arg Ala
                85                  90                  95

Cys Gln Cys Asp Met Cys Arg Leu Ser Arg Tyr
            100                 105


<210> 8
<211> 106
<212> PRT
<213> Human

<400> 8
Gln Glu Ala Val Ile Pro Gly Cys His Leu His Pro Phe Asn Val Thr
  1               5                  10                  15

Val Arg Ser Asp Arg Gln Gly Thr Cys Gln Gly Ser His Val Ala Gln
            20                  25                  30

Ala Cys Val Gly His Cys Glu Ser Ser Ala Phe Pro Ser Arg Tyr Ser
         35                  40                  45

Val Leu Val Ala Ser Gly Tyr Arg His Asn Ile Thr Ser Val Ser Gln
     50                  55                  60

Cys Cys Thr Ile Ser Gly Leu Lys Lys Val Lys Val Gln Leu Gln Cys
 65                  70                  75                  80
```

```
Val Gly Ser Arg Arg Glu Glu Leu Glu Ile Phe Thr Ala Arg Ala Cys
                85                  90                  95

Gln Cys Asp Met Cys Arg Leu Ser Arg Tyr
            100                 105


<210> 9
<211> 105
<212> PRT
<213> Human

<400> 9
Glu Ala Val Ile Pro Gly Cys His Leu His Pro Phe Asn Val Thr Val
  1               5                  10                  15

Arg Ser Asp Arg Gln Gly Thr Cys Gln Gly Ser His Val Ala Gln Ala
            20                  25                  30

Cys Val Gly His Cys Glu Ser Ser Ala Phe Pro Ser Arg Tyr Ser Val
        35                  40                  45

Leu Val Ala Ser Gly Tyr Arg His Asn Ile Thr Ser Val Ser Gln Cys
    50                  55                  60

Cys Thr Ile Ser Gly Leu Lys Lys Val Lys Val Gln Leu Gln Cys Val
65                  70                  75                  80

Gly Ser Arg Arg Glu Glu Leu Glu Ile Phe Thr Ala Arg Ala Cys Gln
                85                  90                  95

Cys Asp Met Cys Arg Leu Ser Arg Tyr
            100                 105


<210> 10
<211> 104
<212> PRT
<213> Human

<400> 10
Ala Val Ile Pro Gly Cys His Leu His Pro Phe Asn Val Thr Val Arg
  1               5                  10                  15

Ser Asp Arg Gln Gly Thr Cys Gln Gly Ser His Val Ala Gln Ala Cys
            20                  25                  30

Val Gly His Cys Glu Ser Ser Ala Phe Pro Ser Arg Tyr Ser Val Leu
        35                  40                  45

Val Ala Ser Gly Tyr Arg His Asn Ile Thr Ser Val Ser Gln Cys Cys
    50                  55                  60

Thr Ile Ser Gly Leu Lys Lys Val Lys Val Gln Leu Gln Cys Val Gly
65                  70                  75                  80
```

Ser Arg Arg Glu Glu Leu Glu Ile Phe Thr Ala Arg Ala Cys Gln Cys
85 90 95

Asp Met Cys Arg Leu Ser Arg Tyr
100

<210> 11
<211> 103
<212> PRT
<213> Human

<400> 11
Val Ile Pro Gly Cys His Leu His Pro Phe Asn Val Thr Val Arg Ser
1 5 10 15

Asp Arg Gln Gly Thr Cys Gln Gly Ser His Val Ala Gln Ala Cys Val
20 25 30

Gly His Cys Glu Ser Ser Ala Phe Pro Ser Arg Tyr Ser Val Leu Val
35 40 45

Ala Ser Gly Tyr Arg His Asn Ile Thr Ser Val Ser Gln Cys Cys Thr
50 55 60

Ile Ser Gly Leu Lys Lys Val Lys Val Gln Leu Gln Cys Val Gly Ser
65 70 75 80

Arg Arg Glu Glu Leu Glu Ile Phe Thr Ala Arg Ala Cys Gln Cys Asp
85 90 95

Met Cys Arg Leu Ser Arg Tyr
100

<210> 12
<211> 102
<212> PRT
<213> Human

<400> 12
Ile Pro Gly Cys His Leu His Pro Phe Asn Val Thr Val Arg Ser Asp
1 5 10 15

Arg Gln Gly Thr Cys Gln Gly Ser His Val Ala Gln Ala Cys Val Gly
20 25 30

His Cys Glu Ser Ser Ala Phe Pro Ser Arg Tyr Ser Val Leu Val Ala
35 40 45

Ser Gly Tyr Arg His Asn Ile Thr Ser Val Ser Gln Cys Cys Thr Ile
50 55 60

Ser Gly Leu Lys Lys Val Lys Val Gln Leu Gln Cys Val Gly Ser Arg
65 70 75 80

Arg Glu Glu Leu Glu Ile Phe Thr Ala Arg Ala Cys Gln Cys Asp Met
85 90 95

Cys Arg Leu Ser Arg Tyr
100

<210> 13
<211> 327
<212> DNA
<213> Human

<400> 13
gcctggggcc aggaggcagt catcccaggc tgccacttgc acccccttcaa tgtgacagtg 60
cgaagtgacc gccaaggcac ctgccagggc tcccacgtgg cacaggcctg tgtgggccac 120
tgtgagtcca gcgccttccc ttctcggtac tctgtgctgg tggccagtgg ttaccgacac 180
aacatcacct ccgtctctca gtgctgcacc atcagtggcc tgaagaaggt caaagtacag 240
ctgcagtgtg tggggagccg gagggaggag ctcgagatct tcacggccag ggcctgccag 300
tgtgacatgt gtcgcctctc tcgctac 327

<210> 14
<211> 324
<212> DNA
<213> Human

<400> 14
tggggccagg aggcagtcat cccaggctgc cacttgcacc ccttcaatgt gacagtgcga 60
agtgaccgcc aaggcacctg ccagggctcc cacgtggcac aggcctgtgt gggccactgt 120
gagtccagcg ccttcccttc tcggtactct gtgctggtgg ccagtggtta ccgacacaac 180
atcacctccg tctctcagtg ctgcaccatc agtggcctga agaaggtcaa agtacagctg 240
cagtgtgtgg ggagccggag ggaggagctc gagatcttca cggccagggc ctgccagtgt 300
gacatgtgtc gcctctctcg ctac 324

<210> 15
<211> 321
<212> DNA
<213> Human

<400> 15
ggccaggagg cagtcatccc aggctgccac ttgcaccct tcaatgtgac agtgcgaagt 60
gaccgccaag gcacctgcca gggctcccac gtggcacagg cctgtgtggg ccactgtgag 120
tccagcgcct tcccttctcg gtactctgtg ctggtggcca gtggttaccg acacaacatc 180
acctccgtct ctcagtgctg caccatcagt ggcctgaaga aggtcaaagt acagctgcag 240
tgtgtgggga gccggaggga ggagctcgag atcttcacgg ccagggcctg ccagtgtgac 300
atgtgtcgcc tctctcgcta c 321

<210> 16
<211> 318
<212> DNA
<213> Human

<400> 16
caggaggcag tcatcccagg ctgccacttg cacccttca atgtgacagt gcgaagtgac 60
cgccaaggca cctgccaggg ctcccacgtg gcacaggcct gtgtgggcca ctgtgagtcc 120
agcgccttcc cttctcggta ctctgtgctg gtggccagtg gttaccgaca caacatcacc 180
tccgtctctc agtgctgcac catcagtggc ctgaagaagg tcaaagtaca gctgcagtgt 240

gtggggagcc ggagggagga gctcgagatc ttcacggcca gggcctgcca gtgtgacatg 300
tgtcgcctct ctcgctac 318

<210> 17
<211> 315
<212> DNA
<213> Human

<400> 17
gaggcagtca tcccaggctg ccacttgcac cccttcaatg tgacagtgcg aagtgaccgc 60
caaggcacct gccagggctc ccacgtggca caggcctgtg tgggccactg tgagtccagc 120
gccttccctt ctcggtactc tgtgctggtg gccagtggtt accgacacaa catcacctcc 180
gtctctcagt gctgcaccat cagtggcctg aagaaggtca aagtacagct gcagtgtgtg 240
gggagccgga gggaggagct cgagatcttc acggccaggg cctgccagtg tgacatgtgt 300
cgcctctctc gctac 315

<210> 18
<211> 312
<212> DNA
<213> Human

<400> 18
gcagtcatcc caggctgcca cttgcacccc ttcaatgtga cagtgcgaag tgaccgccaa 60
ggcacctgcc agggctccca cgtggcacag gcctgtgtgg gccactgtga gtccagcgcc 120
ttcccttctc ggtactctgt gctggtggcc agtggttacc gacacaacat cacctccgtc 180
tctcagtgct gcaccatcag tggcctgaag aaggtcaaag tacagctgca gtgtgtgggg 240
agccggaggg aggagctcga gatcttcacg gccagggcct gccagtgtga catgtgtcgc 300
ctctctcgct ac 312

<210> 19
<211> 309
<212> DNA
<213> Human

<400> 19
gtcatcccag gctgccactt gcacccccttc aatgtgacag tgcgaagtga ccgccaaggc 60
acctgccagg gctcccacgt ggcacaggcc tgtgtgggcc actgtgagtc cagcgccttc 120
ccttctcggt actctgtgct ggtggccagt ggttaccgac acaacatcac ctccgtctct 180
cagtgctgca ccatcagtgg cctgaagaag gtcaaagtac agctgcagtg tgtggggagc 240
cggagggagg agctcgagat cttcacggcc agggcctgcc agtgtgacat gtgtcgcctc 300
tctcgctac 309

<210> 20
<211> 306
<212> DNA
<213> Human

<400> 20
atcccaggct gccacttgca ccccttcaat gtgacagtgc gaagtgaccg ccaaggcacc 60
tgccagggct cccacgtggc acaggcctgt gtgggccact gtgagtccag cgccttccct 120
tctcggtact ctgtgctggt ggccagtggt taccgacaca acatcacctc cgtctctcag 180
tgctgcacca tcagtggcct gaagaaggtc aaagtacagc tgcagtgtgt ggggagccgg 240
agggaggagc tcgagatctt cacggccagg gcctgccagt gtgacatgtg tcgcctctct 300
cgctac 306

<210> 21

<211> 742
<212> DNA
<213> Human

<400> 21
gcaggaggca caggaaaact gcaagccgct ctgttcctgg gcctcggaag tgatgcctat 60
ggcgtcccct caaaccctgg tcctctatct gctggtcctg gcagtcactg aagcctgggg 120
ccaggaggca gtcatcccag gctgccactt gcacccttc aatgtgacag tgcgaagtga 180
ccgccaaggc acctgccagg gctcccacgt ggcacaggcc tgtgtgggcc actgtgagtc 240
cagcgccttc ccttctcggt actctgtgct ggtggccagt ggttaccgac acaacatcac 300
ctccgtctct cagtgctgca ccatcagtgg cctgaagaag gtcaaagtac agctgcagtg 360
tgtggggagc cggagggagg agctcgagat cttcacggcc agggcctgcc agtgtgacat 420
gtgtcgcctc tctcgctact agcccatcct ctcccctcct tcctcccctg ggtcacaggg 480
cttgacattc tggtgggga aacctgtgtt caagattcaa aaactggaag gagctccagc 540
cctgatggtt acttgctatg gaatttttt aaataagggg agggttgttc cagctttgat 600
cctttgtaag attttgtgac tgtcacctga gaagagggga gtttctgctt cttccctgcc 660
tctgcctggc ccttctaaac caatctttca tcattttact tccctctttg cccttacccc 720
taaataaagc aagcagttct tg 742

<210> 22
<211> 601
<212> DNA
<213> Mouse

<400> 22
ggcacgaggg gagtcttcag ttgctgttgg actgtccttt gcagatgccc atggcaccac 60
gagtcttgct cctttgcctg ctgggcctgg cagtcactga agggcatagc ccagagacag 120
ccatcccagg ctgccacttg caccccttca atgtgacggt gcgcagtgat cgcctcggca 180
cttgccaggg ctcccacgtg gcacaggcct gtgtaggaca ctgtgagtct agtgctttcc 240
cttcccggta ctctgtgctg gtggccagtg gctatcggca caacatcacc tcttcctccc 300
agtgctgcac catcagcagc ctcagaaagg tgagggtgtg gctgcagtgc gtggggaacc 360
agcgtgggga gcttgagatc tttactgcaa gggcctgcca gtgtgatatg tgccgtttct 420
cccgctacta gtccccgaag ggctcaggct ccggtcctgc cactgacatg tcatgggtat 480
ctcaaactcg gggctctgac cctctttatc gtctgtgaag atgaggttgg ccctctcagc 540
agtctccttg ctacattctc cttcgctcct gtcctcaata aagcaagcaa tgcttgaaaa 600
a 601

<210> 23
<211> 22
<212> DNA
<213> Mouse

<400> 23
atgcccatgg caccacgagt ct 22

<210> 24
<211> 28
<212> DNA
<213> Mouse

<400> 24
ctagtagcgg gagaaacggc acatatca 28

<210> 25
<211> 22
<212> DNA

<213> Human

<400> 25
gagacatctc cccactgtgt tt 22

<210> 26
<211> 22
<212> DNA
<213> Human

<400> 26
gtttccccca acagaatgtc aa 22

<210> 27
<211> 23
<212> DNA
<213> Human

<400> 27
atgcctatgg cgtcccctca aac 23

<210> 28
<211> 28
<212> DNA
<213> Human

<400> 28
ctagtagcga gagaggcgac acatgtca 28

<210> 29
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: VECTOR

<400> 29
ccatcctaat acgactcact atagggc 27

<210> 30
<211> 23
<212> DNA
<213> Human

<400> 30
gtttccccca acagaatgtc aag 23

<210> 31
<211> 49
<212> DNA
<213> Mouse

<400> 31
aattccacca tgcccatggc accacgagtc ttgctccttt gcctgctgg 49

<210> 32
<211> 40
<212> DNA
<213> Mouse

<400> 32
gcctggcagt cactgaaggg catagcccag agacagccat 40

<210> 33
<211> 26
<212> DNA
<213> Mouse

<400> 33
cccaggctgc cacttgcacc ccttct 26

<210> 34
<211> 45
<212> DNA
<213> Mouse

<400> 34
ctagagaagg ggtgcaagtg gcagcctggg atggctgtct ctggg 45

<210> 35
<211> 40
<212> DNA
<213> Mouse

<400> 35
ctatgccctt cagtgactgc caggcccagc aggcaaagga 40

<210> 36
<211> 30
<212> DNA
<213> Mouse

<400> 36
gcaagactcg tggtgccatg ggcatggtgg 30

<210> 37
<211> 364
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: MOUSE/HUMAN FUSION

<400> 37
Met Pro Met Ala Pro Arg Val Leu Leu Leu Cys Leu Leu Gly Leu Ala
1 5 10 15

Val Thr Glu Gly His Ser Pro Glu Thr Ala Ile Pro Gly Cys His Leu
20 25 30

```
His Pro Phe Ser Arg Glu Ser Thr Lys Asn Ala Leu Glu Thr Trp Gly
      35                  40                  45

Ala Leu Gly Gln Asp Ile Asn Leu Asp Ile Pro Ser Phe Gln Met Ser
  50                  55                  60

Asp Asp Ile Asp Asp Ile Lys Trp Glu Lys Thr Ser Asp Lys Lys Lys
65                  70                  75                  80

Ile Ala Gln Phe Arg Lys Glu Lys Glu Thr Phe Lys Glu Lys Asp Thr
                85                  90                  95

Tyr Lys Leu Phe Lys Asn Gly Thr Leu Lys Ile Lys His Leu Lys Thr
            100                 105                 110

Asp Asp Gln Asp Ile Tyr Lys Val Ser Ile Tyr Asp Thr Lys Gly Lys
        115                 120                 125

Asn Val Leu Glu Lys Ile Phe Asp Leu Lys Ile Gln Glu Arg Val Ser
    130                 135                 140

Lys Pro Lys Ile Ser Trp Thr Cys Ile Asn Thr Thr Leu Thr Cys Glu
145                 150                 155                 160

Val Met Asn Gly Thr Asp Pro Glu Leu Asn Leu Tyr Gln Asp Gly Lys
                165                 170                 175

His Leu Lys Leu Ser Gln Arg Val Ile Thr His Lys Trp Thr Thr Ser
            180                 185                 190

Leu Ser Ala Lys Phe Lys Cys Thr Ala Gly Asn Lys Val Ser Lys Glu
        195                 200                 205

Ser Ser Val Glu Pro Val Ser Cys Pro Glu Lys Gly Leu Asp Ile Tyr
    210                 215                 220

Leu Ile Ile Gly Ile Cys Gly Gly Gly Ser Leu Leu Met Val Phe Val
225                 230                 235                 240

Ala Leu Leu Val Phe Tyr Ile Thr Lys Arg Lys Lys Gln Arg Ser Arg
                245                 250                 255

Arg Asn Asp Glu Glu Leu Glu Thr Arg Ala His Arg Val Ala Thr Glu
            260                 265                 270

Glu Arg Gly Arg Lys Pro Gln Gln Ile Pro Ala Ser Thr Pro Gln Asn
        275                 280                 285

Pro Ala Thr Ser Gln His Pro Pro Pro Pro Pro Gly His Arg Ser Gln
    290                 295                 300

Ala Pro Ser His Arg Pro Pro Pro Pro Gly His Arg Val Gln His Gln
305                 310                 315                 320

Pro Gln Lys Arg Pro Pro Ala Pro Ser Gly Thr Gln Val His Gln Gln
                325                 330                 335
```

```
Lys Gly Pro Pro Leu Pro Arg Pro Arg Val Gln Pro Lys Pro Pro His
        340                 345                 350

Gly Ala Ala Glu Asn Ser Leu Ser Pro Ser Ser Asn
    355                 360
```

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 00/16541

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/16 C12N15/62 C12N5/10 C07K14/59 C07K16/26
A61K38/22 A01K67/027 G01N33/50

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K A61K A01K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, STRAND

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 99 41377 A (ZYMOGENETICS INC) 19 August 1999 (1999-08-19) abstract page 30, line 20 - line 21 page 31, line 3 - line 6 page 32, line 18 -page 33, line 4 examples 1-14 sequence of zSIG51 (seq. ID 2). claims 1-28	1-12, 17-46
X	DATABASE EMBL SEQUENCES 'Online! Accession No. AI693322, 3 June 1999 (1999-06-03) STRAUSBER R.: "EST; H. sapiens cDNA clone IMAGE:2338950" XP002148574 the complementary strand of nt 309-521 encodes aa 58-129 of the protein claimed.	1-12, 17-21

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 September 2000	10/10/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Galli, I

C.(Continuation) DOCUMENTS CONSIDERED T BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	SUN ET AL: "the cysteine-knot growth factor superfamily" ANNUAL REVIEW OF BIOPHYSICS AND BIOMOLECULAR STRUCTURE,ANNUAL REVIEWS INC., PALO ALTO, CA,US, vol. 24, 1995, pages 269-292, XP002104603 ISSN: 1056-8700 the whole document	1-12, 17-46

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 13-16 and partly 17

Claims 31,32,37-40 and 44 refer to a "nucleic acid of claim 3", but claim 3 relates to a polypeptide. These claims probably ought to refer to the nucleic acid of claim 5. The earch has been performed on the basis of this assumption.

Claims 13-16, and partly claim 17, refer to processes to produce the polypeptide claimed by "genetically engineering a mammalian cell to express the polypeptide from one or more of the cell's endogenous genes". The scope of these claims is unclear, as they relate in very general terms to results to be achieved and without offering a true technical characterization of any kind. Furthermore, no such processes are described in the application. These claims are therefore ambiguous and vague, and the subject matter is not sufficiently disclosed and supported according to Art. 5 and 6 PCT, to such extent that a meaningful search is impossible.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 00/16541

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9941377 A	19-08-1999	AU 2676399 A	30-08-1999